Stress and Salience in English: Theory and Practice

By

Henry Swift Thompson

A.B. (University of California) 1972
M.S. (University of California) 1974
M.A. (University of California) 1977

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Linguistics

in the

GRADUATE DIVISION

OF THE

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

5/8/80
Chairman Date

Charles J. Fillmore 5/8/80

Stephen E. Palmer 5/12/80

DOCTORAL DEGREE CONFERRED
JUNE 14, 1980

Table of Contents

Acknowledgments

This work would not have been possible without the facilities of the Computer Science Laboratory of the Xerox Palo Alto Research Center and the support of its members. In particular I would like to thank Danny Bobrow, Jerry Elkind, Martin Kay and Bob Taylor for helping to provide the resources and the environment, both material and intellectual, which enabled me to do that which I report herein and much more besides.

My thanks also go to the faculty and to my fellow students in the Department of Linguistics at the University of California, Berkeley. My committee, George Lakoff, Charles Fillmore, and Steve Palmer, provided support and useful criticism in appropriate measure, and I am grateful for both of them. Deborah Davisson, John Kingston, Cathy O'Connor and Ken Whistler have all lent a helpful ear to various parts of this work, and for this and much support in general I am in their debt.

The care, concern, and support of many friends, not themselves linguists but sensible nonetheless of the pressures accompanying the writing of a dissertation, have been a great blessing in a time of trial. To Arlene Baxter, Ken Pier, Willie-Sue Haugeland, and Tom and Barbara Welch, then, my love, as always.

Many of the ideas contained herein first saw the light of day in discussions with Mark Liberman, and for his interest and helpful criticism I am most grateful.

And finally, heartfelt thanks to Dianne Morton, to whose long and uncomplaining labors is owed the polished appearance of this document.

The work reported herein was supported in part by a National Science Foundation Graduate Fellowship, and in part by a Xerox Corporation Research Internship.

Chapter 0. Introduction

Consisting of a motivational and historical sketch in four parts, looking both backward and forward, in which the reader is introduced (!) to the problem.

0.1 The motivation for the inquiry - variation in the form of expression

One of the major unsolved problems in theoretical linguistics is that posed by the multitude of different syntactic forms available for the expression of truth-conditionally equivalent sentences. Associated with a simple active sentence are the passive, dative shifted, topicalized, it-clefted, wh-clefted, PP-preposed, and so on, versions, which present more or less the same constituents rearranged in various ways, with little if any effect on propositional content. For many years the emphasis within linguistics with respect to this problem has been on the complete and accurate description and specification of the alternate forms, but more recently the question of *why* this multiplicity of forms is provided in the language has begun to receive more attention.

0.1.1 Realistic/functional approaches to grammar

In particular there has been a growing trend in linguistics recently towards what Nichols [1979] has called *strong* theories of grammar, that is, theories that are understandable not only as descriptions of linguistic regularity, but also as having direct reflections in the psychological mechanisms which participate in human linguistic behavior. This trend is exemplified in e.g. [Lakoff and Thompson 1975a, 1975b], [Thompson 1976, 1977], [Bresnan 1978], [Kaplan 1975] and [Frazier and Fodor 1978]. Although presenting considerably different proposals in detail, these and other similar works share a common commitment. It is that linguistic theories should be judged not only on how well they can describe the data, but also on whether or not a plausible processing model can be formulated which incorporates them. This bears directly on the question posed above, for a processing model needs must address the issue of what conditions the choice among the various forms available within the language for expressing a proposition. It is not explanatory in an important sense unless it does so. That is, a processing model which is capable of producing the range of English sentences in a psychologically plausible way, starting with some sort of psychologically plausible underlying form, is nonetheless not fully satisfactory if that underlying form encodes in some overt way the distinctions between syntactic alternatives.

If the underlying form simply marks some incipient noun phrase as 'topic', or some underlying predicate as 'to be passivized', then an important linguistic issue has not been addressed. The question of what it is that conditions the alternation between the active and passive versions of a sentence, or the preposing of a prepositional phrase, cannot be ignored. A solution to that problem, hereafter referred to as the *conditioning problem*, would in turn go a long way toward answering the question of why the different forms exist, by providing a functional basis for differentiation among them.[1]

The beginnings of an answer have been around for some time. The various forms may be truth-conditionally equivalent, but there are strong constraints on their appropriateness. Thus in the case of double object sentences, for example, the unexceptionable question-answer pairs in (la) and (lb) and the aberrant pairs in (2a) and (2b) below illustrate the conditioning of the two possible versions by the form of the preceding questions.

la) *What did John give to Bill?*
John gave Bill the book.

lb) *To whom did John give the book?*
John gave the book to Bill.

2a) *What did John give to Bill?*
**John gave the book to Bill.*

2b) *To whom did John give the book?*
**John gave Bill the book.*

Among others, Michael Halliday [1967a, 1967b] and Petr Sgall [Sgall, Hajičová, and Benešová 1973] have both articulated relatively detailed theories which attempt to describe and explain this and other similar phenomena which are in the domain of the conditioning problem in terms of the information structure of the surrounding discourse, appealing to concepts like *given* vs. *new, topic* vs. *comment* or *focus*, and *theme* vs. *rheme.* What is significant about their approach is that the domain of investigation has been widened. To account for variation at the sentential level, appeal has been made to larger units of linguistic structure. Indeed, if there is anything in common among the extant proposals concerning the conditioning problem, it is that the domain of concern is the 'discourse,' construed variously, but larger than the sentence at any rate.

[1] This is not to imply an insistence that *all* variation be accounted for by functional conditioning - there will surely be some residue which is essentially capricious, although as Halliday points out (personal communication), appeal to caprice or style is just a formulaic admission of defeat.

0.1.2 Syntax above the sentence

This expansion of concern to the discourse level intersects a number of other trends both within and without linguistics. There is a long standing practice in descriptive field linguistics of including the analysis of texts and the textual function of words and constructions in grammatical descriptions of languages. These texts are typically stories or so-called ethnographic texts, which is to say descriptions of everyday life, and are almost exclusively oral. Linguistics theories which have developed from a concern with field studies, e.g. tagmemic grammar ([Longacre 1977]), have from the beginning been concerned with grammar at the textual level. From another direction, European proponents of various forms of 'text grammar', (e.g. van Dijk [1972], Petofi [1973], even Propp [1968]) have also tried to apply more or less traditional structural analysis to texts. Starting from concerns similar to those outlined above, psycho-linguists (e.g. Clark and Clark [1977], Haviland and Clark [1974], Kintsch [1974]), linguists (Chafe [1974, 1976], Grimes [1972], Givón [1976]), sociolinguists (e.g. Labov and Fanshell [1977], Linde [1974]) in this country have been increasingly concerned with discourse level phenomena.

But there is a confusion at the base of this inquiry, which has significantly impeded progress. It is a confusion which results from the mis-application of the structural-syntactic paradigm. Jerry Morgan [1978] points out that it is not really appropriate to talk about 'syntax' at any level about the sentence. There are very few cases where phenomena which are truly syntactic occur above the sentence level. That is, there are few if any cases where the linguistic *form* of one sentence or part thereof has a direct effect on the *form* of another sentence or sub-part. The relationships which do exist between sentences seem better viewed as being mediated through content. The problem seems not to be one of understanding the relationship between prior text and subsequent text, but rather of that between text and context, both in terms of how the context at any point is partially determined by prior discourse, and of how the context affects the expression of the text at any given point.

Thus the relationship exemplified in examples (1a) and (1b) above is not a direct one between the form of the question and the form of the response, but rather an indirect one between the context established by the content of the question and the form of the response. The 'illformedness' of (2a) and (2b) is not syntactic but rather semantic/pragmatic. They don't seem to make sense. They are discordant in the same way that (3) and (4) below are:

3) *Morris is bald.*
(Same speaker) *Since Morris is not bald ...*

4) *I don't own a chihuahua.*
(Different speaker) *I regret that my cobra bit your chihuahua.*

If someone actually responded as in (2a), he would be assumed to have misunderstood the question, rather than being accused of a lack of knowledge of English. And if it turned out that he did in fact have a linguistic deficit, it would be located at the sentential level. It is not the *linguistic* context, in this case *What did John give to Bill?,* which conditions the form of the response. Rather the form of the response is appropriate only to a certain class of cognitive and real world contexts, and the question establishes a context inconsistent with that class.

The decrease in applicability of grammaticality judgments as the size of the linguistic unit judged increases supports this distinction between direct, syntactic relationships and indirect, semantic/pragmatic ones. There does not seem to be any consistent, non-trivial interpretation of the notion of grammaticality which allows both (2a) and (2b) on the one hand and (5a) and (5b) on the other to be judged ungrammatical on the same, or anything but metaphorically related, grounds.

5a) **green a trees tall* (as a noun phrase)
5b) **did running been have* (as a verb group)

The *s on (2a) and (2b) seem much more to reflect some sense of interpretational mismatch or failure of communication, whereas (5a) and (5b) are unarguably syntactic. Judgments about sentences seem to fall somewhere in between, with some (e.g. (6)) seeming quite syntactic and clear, and others (e.g. (7a) and (7b)) much more contextual and/or pragmatic.

6) **Those animals am a horse*

7a) **Julian greeted everyone when Sandy greeted everyone*
(with a distributed reading) [Sag and Weisler 1979]

7b) **We found a person whose car to drive to the meeting* [Baker 1979]

A considerable amount has been written on the issue of the value of grammaticality judgments even at the sentential level, but even with the outcome of that debate in doubt it seems clear that they are not plausible at any level above the sentence, precisely because of the qualitatively different nature of inter- as opposed to intra- sentential relationships.

Thus 'syntax above the sentence' is really a misnomer. What is really at issue is the contextual dependency of linguistic forms, both lexical and syntactic. Texts do not 'have' information structures or cohesion - rather a 'good' text is the simultaneous articulation and maintenance of a context consistent with all of the utterances which compose it. It is hard to see how significant progress can be made either on problems of discourse structure in general, or on the conditioning problem in particular, without recognizing and profiting from the difference between this approach and the structural-syntactic one. Unfortunately, almost all of the powerful and insightful tools and methodologies of linguistics have been in the structural domain, and it in part for lack of an alternative that they have been applied to discourse despite their inappropriateness for the task.

0.1.3 Models of linguistic processing

One alternative which has been adopted in many cases is to make proposals informally in (con)textual or psychological terms - 'The given/new contract', 'Functional Sentence Perspective', 'Communicative Dynamism', and so forth. Some specification of contexts and the ways in which they may condition syntactic form have been attempted, almost exclusively in anecdotal terms. But because of their informal, anecdotal nature such proposals have often had difficulty in being taken seriously, as they seem to make no verifiable predictions or falsifiable claims.

It is here that the development of processing models of language production and comprehension could be helpful. Within the context of a production model, one might hope to be able to provide explicit definitions of terms like *given/new* or *theme/rheme* in terms of the structures and processes of which such a model is composed. Process modelling also improves the ability to verify theories, in that if a sufficiently comprehensive model is available, its predictions about the form of a particular utterance can be compared with actually occurring 'live' data. More importantly, in terms of the distinction discussed above with respect to discourse, the processing model approach seems well suited as a way of thinking about the problem, encouraging rather than discouraging as it does a focus on the role of context in mediating between the content of prior text and the form of subsequent text.

My interest has been for some time in the production aspect of process modeling. An obvious source of motivating data for the development of such a model is natural, unselfconscious, single speaker speech. This is a good place to begin for two reasons. First, it is full of incidental cues to the production process which are rare or absent in more self-conscious forms: filled and unfilled pauses, false starts, corrections, and so on. Second, its very unselfconsciousness suggests that it is a relatively uncontaminated form of linguistic behavior. In other forms, conscious intrusions may affect the process and the result. The form of such intrusion is of course worthy of study, but needs must be based on a model of the basic unmodulated process. Written text is especially suspect in this regard. The extent to which sentences in written texts take the form they do in response to the differences in medium (impoverishment, with respect to prosody and interaction, enrichment, with respect to permanence. See [Rubin 1978] for discussion) is poorly understood at best.

In fact, given that the goal is to investigate the contextual and psychological conditioning of linguistic form, and the language concerned is English, the impoverishment of written text with respect to prosody is critical. Where many languages use some form of lexical or inflectional system to encode speaker attitude and informational status, such as particles or aspect markers, English uses prosody. The written form of English thus is missing vital cues as to how the speaker viewed the context which are not missing from the written form of e.g. Cantonese, which uses particles in many cases where English uses prosody. For me, the task of understanding how to extract and categorize the contribution of prosody to the background of the conditioning problem came dominate the earlier goal of investigating the problem itself, and forms the basis for this thesis.

0.2 The problem of prosody

0.2.1 Live data makes prosody an issue

For all of this has been not only an introduction, but also a historical recapitulation. I originally proposed to write a thesis on *Factors Conditioning Word Order in Spoken English Clauses.* Some time ago, my interest in language production models led me down the path described above, to the point of trying to examine some 'live' data with a view to characterizing the information structure that would condition, for a simple production model, the constituent order observed in the text. This necessitated venturing forth with a tape recorder to collect some data, which was not too difficult, and then transcribing it in preparation for subsequent analysis.

At this point the issue of prosody[2] intruded. For as should be clear from examples (1) and (2) above (repeated below), prosody is crucially involved in this enterprise. The responses in (2) are only aberrant in their spoken manifestations, spoken with the same prosody as in (1), as answers to the opposite question. In fact it does not really make sense to judge the written forms at all, for they can only be called ill-formed with respect to the assumption of a complex and as yet unsubstantiated set of assertions concerning the way in which people impute prosody to sentences as they read them. On the other hand, if they are spoken as in (8) below, with strong accent on the indicated words, they are perfectly acceptable.

1a) *What did John give to Bill?*
John gave Bill the book.

1b) *To whom did John give the book?*
John gave the book to Bill.

2a) *What did John give to Bill?*
**John gave the book to Bill.*

2b) *To whom did John give the book?*
**John gave Bill the book.*

8a) *What did John give to Bill?*
*John gave the **book** to Bill.*

8b) *To whom did John give the book?*
*John gave **Bill** the book.*

It seems that to the extent that we can characterize the contextual implications of any constituent order variation, there is some prosodic variation which has similar if not identical implications. To say nothing of the interesting second order question this raises concerning the conditioning of the choice between 'using' prosodic versus constituent order variation in a particular situation, e.g. (1a) vs. (8a), this makes the task outlined above much more difficult in an intensely practical way. In general, as pointed out at the end of the previous section, prosody plays a pervasive role in English at the discourse level. It is no longer sufficient to just transcribe the words of the data. The prosody must be transcribed as well. And the sad truth

[2] A brief note on terminology is in order: I use *prosody* as a cover term for supersegmental phenomena of all sorts, with *salience* and *intonation* as the two principal perceptual prosodic sub-categories. Following Bolinger, I reserve *stress* for an abstract property, the potential locus of salience, whether within lexical items, phrases, or larger units. More precise definitions will be presented in chapter 1 below.

is that although it is easy to tell whether *Bill* precedes or follows *the book*, it is not easy at all to tell whether a word is accented strongly or not.

But there is something encouraging about this as well. *Prosody* is a cover term for a complex system of linguistic phenomena. It has syntagmatic and paradigmatic structure of its own, at least partially independent from the more obvious syntactic structures it co-occurs with. Most importantly, it seems that in attempting to uncover and explicate the function of prosodic phenomena, we are led in the same direction as when we seek an answer to the conditioning problem, as discussed above. Thus not only is some at least taxonomic understanding of prosody a prerequisite for using live data to investigate the conditioning problem, but also one may hope for a synergistic coincidence between the theoretical apparatuses which the two problems require.

Nevertheless before that hope can be realized the level of descriptive, taxonomic adequacy must first be achieved. Unfortunately despite a large amount of work that has been done, there is far from any consensus as to the correct approach to English prosodic phenomena, either at the theoretical level or at the level of notation and transcription methodology. This is not to suggest that no work of value has been done in the field. My proposals are based on positions first propounded by Abercrombie [1964], Halliday [1967a, 1967b], Bolinger [1958, 1965a, 1965b, 1972], and more recently, Liberman [1975], Liberman and Prince [1977], Pierrehumbert [1979, 1980], Selkirk [1979, forthcoming], and Ladd [1978]. Indeed both Liberman and Ladd have made significant progress towards a prosodic theory which would be an adequate basis for syntactic investigations at the discourse level, but that goal is still unreached, and it is the purpose of this thesis to advance further along the way. What follows is an attempt to make some progress on both the practical and the theoretical fronts, trying to move toward a transcription methodology and a theoretical framework which would support the kind of project I had originally had in mind. The methodological position outlined at the beginning of this introduction, with its commitment to the use of live data and the need for psychologically plausible theories, has informed both efforts.

0.2.2 The problem of stress and salience

There is one sub-part of the prosodic problem which is of considerable importance, yet has received little attention. Whereas on the one hand a great deal has been written about intonation, and considerable controversy exists over approaches to the notation of tunes, issues of intonational meaning, and so on, and on the other there is a definite tradition of study of stress at the word level, nonetheless little if anything convincing has been written about stress and salience above the word level. The time is right to try to fill this gap. Starting from Liberman's new approach to word stress, and building on his reappraisal of Abercrombie's and Halliday's emphasis on the rhythmic basis of stress and salience phenomena, it is my purpose in this thesis to demonstrate that a consistent approach to stress and salience at all levels is now possible, an approach which by and large decouples salience phenomena from intonational phenomena, an approach which is grounded in live data and yields psychologically plausible results.

0.3 Things to come

In summary, my goal is to present a theoretical account of some aspects of English prosody, in particular of the factors involved in the relationship of abstract stress and manifest salience. This account is formulated as a processing model of the relevant aspects of the speech production process, for reasons of methodological preference on my part. The practical foundation for this theoretical activity is provided by the controlled application of a transcription methodology to natural speech.

The balance of this work is organized as follows: Chapter 1 provides a synthesis of past and present approaches to the problem, to serve as a background against which my practical and theoretical contributions are set. Chapter 2 presents a transcription notation and methodology grounded in live data, and the results of an experiment designed to test them. Chapter 3 presents some measurements of aspects of the text used in that experiment, based on the consensus transcription. Chapter 4 presents the production model which embodies my theoretical claims. Chapter 5 applies this model to the consensus transcription and sums up, and appendixes follow with a glossary, some of the original experimental data, and a brief description of the various computer systems which I developed and/or used to assist in the gathering and processing of that data.

Chapter 1. Prosody: An introduction to the terminology and the problems

Consisting of an introduction to the prosodic phenomena of English, together with definitions of the terms I will employ, in which a distinction is made between two aspects of prosody, *organizational/temporal* and *categorial/tonal*, followed by the posing of a number of questions to be addressed in the balance of the thesis, with a brief survey of the treatments they have received from other workers in the field.

1.1 Prosodic phenomena and terminology

Any introduction such as this faces serious organizational problems. There is little agreement in the field on any but the most trivial aspects of the phenomena under study, and the terminology in use is inconsistent and frequently contradictory. Any presentation is bound to project a set of methodological and theoretical prejudices which some will find obvious, others controversial, and still others patently absurd. One possible way to reduce this problem is to exhaustively survey the relevant literature before presenting my own position. I do not propose to do this, because it has been done better than I could already [Crystal 1969], and because it is my intention in this thesis not to engage in extensive, point by point comparisons of my proposals with others. Rather I propose to set out what I think is a coherent approach, whose principal justification for my readers must come from their independent judgment of its consistency and utility in confronting the data they consider relevant. Thus what follows will be presented as if I were the confident possessor of the truth, setting it forth as directly as may be. Major debts will be acknowledged, rank speculation will occasionally be identified as such, and points of major controversy will be noted, but extensive arguments contra other proposals will not be found.

1.1.1 Prosody itself

Prosody is one of a class of words we use uncritically to refer to aspects of linguistic phenomena. It is used in the same way as e.g. *case*, or *deixis*. But the status of such words is somewhat obscure. The exposition of terminology below will be made easier if their dichotomous nature is clarified. From the perspective of the analyst, these words refer to a collection of characteristics of utterances which he takes to be a coherent system. That is, they seem more easily explained if treated as related than if treated as not related. The existence of a recognizable parallel between two paradigmatic sets, one phenomenal and the other functional, is usually taken as grounds for positing and naming such a linguistic system. Thus *deixis* is founded on the apparent relation between the phenomenal paradigm made up of the words *this* and *that* and the functional paradigm of the location of referents as proximal or distal.[1] Such words may be used to refer to either the phenomenal or the functional aspects of the system, or of the systematic conjunction of the two. In the definitions which follow, I will usually try to present both the phenomenal and the functional attributes of a word. Our understanding usually starts with the phenomenal aspect, since it is the more accessible to relatively objective description, usually in acoustic terms. The delineation of the corresponding functional paradigm is often less clear, and in fact may constitute the theoretical question of concern.

From the phenomenal perspective, the following non-exhaustive list suggests the sort of acoustic properties of an utterance which are called *prosodic*: amplitude, fundamental frequency, timbre or tone of voice, phonation type (normal, whisper, breathy voice), duration and timing. From the functional perspective, the prosodic aspects of an utterance are those aspects of its acoustic form which are non-segmental. By non-segmental I mean loosely to select those acoustic properties of an utterance which are *not* strongly determined by its morphological constituency. These two definitions are clearly not completely equivalent, as morphological choices may in some cases determine the properties listed above either directly, as in a language with lexical tone, or indirectly, as for instance when the number of syllables in an utterance restricts its possible pattern in time. Nonetheless all the items on the foregoing list are at least partially independent of morphological determination, and the obvious synthesis is to say that for any given utterance, its prosodic aspects are as listed above, *to the extent* that they are not morphologically determined. The name for the part of the grammar which concerns itself with such aspects is *prosody*. Put another way, when the morphemes of an utterance have been specified together with their order, this typically does not uniquely determine an acoustic form for the utterance. What remains unspecified is prosodic, and the domain of prosody is the study of the constraints a language imposes on that remaining determination.

[1] As Jespersen [1924] points out, there is yet another paradigm, which he calls *notional*, which is related to the functional. It is 'in the world', and thus for deixis we have e.g. closer to/further from the speaker or after/before the current utterance mapping onto proximal/distal.

Not all the aspects of prosody so defined are of concern to us here. In particular I shall have nothing to say about timbre, tone of voice, or phonation type. They function almost exclusively at the inter-personal or affective level of linguistic interchange, and hence are extremely difficult to pin down as to specific functional role.

1.1.2 Organizational/temporal and categorial/tonal

I believe that from the functional perspective prosodic phenomena can mostly be assigned to one of two distinct sub-systems - *organizational*, which imposes a structure on utterances above the word level which reflects the speaker's structuring of the information being communicated, and *categorial*, which categorizes elements of an utterance in various ways which reflect aspects of the speaker's attitude. One of my goals in this thesis is to define these terms so that their referents are clear and the assertion that those referents are distinct is established. From the phenomenal perspective this distinction between organizational and categorial correlates loosely with a distinction between *temporal*, which covers the durational and rhythmic and timing aspects of the acoustic form of an utterance, and *tonal*, which covers the aspects of fundamental frequency and, to a lesser extent, amplitude.

Strict adherence to the functional/phenomenal dichotomy will not be possible in the exposition that follows. Some prosodic phenomena are obviously manifested phenomenally, while their functional specification is unclear. Others have a clear functional role, but lack a clearly stateable phenomenal exponent. Therefore in what follows I will sometimes adopt one stance, sometimes the other, as appropriate to the particular prosodic phenomenon being introduced. There is a circular dependency in the definitions and partitions I impose, both from one to another and from the phenomenal aspect to the functional and back again, which cannot be avoided. "Circular explanations for circular facts", as Martin Kay has put it, but of course the problem is that, at least in this case, there is no 'fact of the matter', at least not in any currently accessible form, merely a mutually determining set of perceptions and interpretations. Picking apart such a circle to lay it out linearly on paper inevitably fails at the start, but hopefully by the end the reader can rejoin the circle and see that in fact it all does fit together. Thus if I appear to start in the middle, it is because there is nowhere else available.

1.1.2.1 Organizational/temporal

The flow of speech is organized into a hierarchical structure of units in time. From small to large, these units are the syllable, the word, the foot, and the tone group. The syllable and the word are not strictly speaking prosodic units, and will not be dealt with here. The foot and the tone group are the two main organizational/temporal prosodic phenomena.

1.1.2.1.1 The ***tone group***

The larger of the two, and the largest clearly identified prosodic unit, is the *tone group*.[2] People perceive connected speech as divided into stretches which have a rhythmic and tonal coherence, with perceptible boundaries between each such stretch. *Tone group* is the name given to this unit from the phenomenal perspective. Its boundaries are accompanied by various combinations of pause, tempo change, and pitch movement. Although the exact acoustic and perceptual correlates of the boundary between tone groups is far from clear, nonetheless people's perceptions of their location in utterances are confident and stable (see section 2.2.3 below). It is the domain of the tonal phenomena described below, and it is in large part their distribution within a tone group which gives that tone group its own identity.

In a text sufficiently well controlled so as to consist largely of complete clauses, tone groups and clauses will frequently be co-extensive, but this is not necessarily the case. Frequently a clause will be covered by several tone groups, co-extensive with noun or verb groups, or even single words. Sometimes a single tone group will span a multi-clause sentence, or even several sentences.[3]

[2] The origin of the technical terms defined here and elsewhere, together with a brief definition and reference to their full definition in the text, can be found in the glossary in Appendix A.

[3] Some portions of utterances are not part of a tone group, or are at best defectively or trivially tone groups unto themselves - words such as *and* or *so* and filled pauses, when they occur in isolation and serve only to hold the floor.

*1.1.2.1.2 The **information unit***

From the functional perspective, this largest prosodic unit is called the *information unit.* Its exact characterization from this stance is far from clear. Intuitively .it serves as a single communicative block, possibly representing a level of structure in the production process - the encoding of one coherent thought. It is the domain of the categorizations provided by the categorial structure described below, and as above for the tone group, it is the articulation of these categories across a recognizably coherent domain that lends credence to the identification of that domain as an organizational unit.

There has been some suggestion of a unit larger than the tone group, which would correlate more nearly with the sentence, sometimes on phenomenal grounds (e.g. by Crystal [1969], Halliday [1967a], and Liberman [1975]) and sometimes on functional ones (e.g. by El-Menoufy [1969] and especially Chafe [1977, 1979a]), but that is beyond the scope of my concerns here.

*1.1.2.1.3 The **foot***

The the foot is fundamental rhythmic unit of speech. The rhythmic quality of speech is most clearly felt in poetry, where we have a distinct experience of regularly recurring patterns, as in the limerick below.

I knew a young lady named Bright
Who traveled much faster than light
1) *She departed one day*
In a relative way
And arrived the preceding night

Just as in music the rhythmic division into measures results in a distinction between strong and weak, so in speech the division into feet results in a distinction between *salient* and *non-salient* syllables. Thus in (1) above in each of the first, second, and last lines there are three feet, and three salient syllables, while in the third and fourth lines there are only two feet, and two salient syllables. A foot consists of one or more syllables, and it is the first syllable of each foot which is the salient syllable. The first syllable is salient with respect to the other syllables in the foot, and they are non-salient with respect to it, and it is this relative, relational interdependence which gives the foot its identity.

The acoustic correlates of these relations are unclear, being some combination of length, pitch, and loudness, approximately in that order. The salient syllable is likely to be some combination of longer, higher in pitch, and louder than the non-salient syllables in a foot. For a discussion of this see e.g. [Fry 1955, 1958], [Lea 1977], [Lehiste 1973], and [O'Malley et al. 1973]. It is important to understand that for me (following Halliday and Liberman) the foot is primarily a rhythmic unit, and salience is a derivative notion, dependent on the foot. Thus it is not that we hear some syllables as salient, and from that determine foot boundaries, but rather that we hear feet, and and as a result determine some syllables to be salient. In come cases a foot may be perceived for rhythmic reasons alone, without any explicit acoustic cues.

Ladd [1978] presents an excellent survey of the rhythmic approach to salience, including the following quote from Householder [1957], who presents very well the concept of a rhythmic structure partially guided but not totally determined by acoustic cues: "The fact is we can't hear noises repeated with fair regularity at more than a certain average frequency without grouping them rhythmically (as every subway-rider can testify), and once a given pattern is established we will hear it over and over till some new irregularity breaks the rhythm and starts another pattern." Some interesting experimental support for this position is given in [Bell 1977].

The words *stress*, *accent*, and *salience* have been used in many different ways by different authors in the field, and it is important to be clear about the exact way I am using them. *Salience* is a relational, phenomenal property of syllables, derivative from the phenomenal, rhythmic unit, the foot. I follow Bolinger (and few others) in reserving *stress* for an abstract property, the potential locus of *salience*, which is what is actually manifest in an utterance. I will not use *accent* at all.[4] It follows that an unstressed syllable may be salient. For example the word *emphasis*, which has in the abstract e.g. eight letters and three syllables, also has the abstract property of stress on its first syllable, whereas a particular token of that word, as might occur in the gag line *he always puts the emphasis on the wrong syllable*, has the concrete property of salience on its second syllable. Likewise, in the last line of the limerick given above as (1), for rhythmic reasons the first syllable of *preceding* is salient, although in the abstract it is the second syllable which is stressed.

Stress has a confused history of use. For the American Structuralists it was a perceptual property of syllables. There were some fixed number of absolute levels of stress, with the strongest

[4] Note that this is in almost diametric opposition to that of e.g. Halliday, who uses *accented* where I will use *stressed*. In common parlance one also often finds *stressed* used where I will use *salient*.

(*sentence stress*) restricted to occurring once per sentence. There was some disagreement as to whether a reduced syllable was assigned the weakest stress and that was it, or whether it had an additional feature. It was apparently Daniel Jones who introduced the analysis in terms of four levels to America, although he later changed his mind.

Chomsky and Halle in *The Sound Pattern of English* [1968] adopted this view into the generative school, but made things somewhat more abstract by making the number of levels unbounded. They retained the notion of sentence stress, or 1 stress, and promulgated three rules, the *English Stress Rule*, which assigned stress levels to the syllables of individual words based on their phonemic structure, the *Compound Stress Rule*, which adjusted the stress levels of words joined together into compounds, and the *Nuclear Stress Rule*, which further adjusted the stress levels of words and compounds combined into sentences.

The British school in its various versions typically distinguished only stressed, unstressed and reduced, as they did not consider what the Americans called *sentence stress* to be stress at all.

Halliday, following Abercrombie, viewed what he called salience as a rhythmic phenomenon, and did not use the word stress at all.

Bolinger introduced a much needed distinction between the abstract notion of (lexical) stress, and the actual occurrence of what he called accent, which is closer to what I will call tonal excursion below than it is to salience, as it is defined strictly in terms of pitch movement.

Liberman and Ladd, starting from the rhythmic view of things, developed an approach to stress and salience which is uniform for both words and sentences, although they did not use the words as I do. This uniformity raises the issue of whether the distinction between abstract and concrete can be profitably extended to the sentence level.

As laid out above, my use of the word *stress* follows Bolinger's. In particular, I do not make the extension to the sentence level proposed by Liberman and Ladd: for me stress is a property exclusively of (poly-syllabic) lexical items.

All of the above discussion of feet, stress, and salience has been from the phenomenal perspective. The functional characterization of feet is less clear, and indeed they may not need any which is particular to language, in that rhythmic structure is something people impose on repetitive occurrences in time, regardless of the perceptual modality involved. The functional explanation for this in turn is also unclear, but is not (solely) the responsibility of linguistics. It is attractive to speculate that in fact the acoustic differentiation of salient versus non-salient syllables may have depended for its development on the independently established rhythmic structure, and not the other way around. As speech speeds up and various distinctions neutralize under pressure to communicate faster and with less attention, retention of accurate articulation (salient syllables remaining longer and louder than their reduced counterparts) at *independently predictable intervais* is a good compromise. It would function to aid comprehension, by providing an independent, rhythmically based cue to islands of acoustic reliability, which in turn provide a stable nucleus from which syllable and word finding processes can work outward, if the insights gained in the attempts at speech recognition by machine have any applicability to the process in the human brain (see e.g. [Allen, Jonathan 1978]).

1.1.2.1.4 The relation of foot to tone group, and notation

Tone groups are composed of feet, possibly preceded by one or more syllables of upbeat, which, like the non-salient syllables of feet, are perceived as rhythmically weak, subordinate to the salient syllable at the head of the following foot. In a typical way of reciting the limerick given above as (1), the structure at the foot and tone group level would be as follows:

| *I / knew a young / lady named / Bright* |
Who / traveled much / faster than / light |
2) *She de / parted one / day* |
In a / relative / way |
And ar / rived the / preceding / night |

where vertical bars (|) are used to mark tone group boundaries, and slashes (/) to mark foot boundaries. By convention the slash which would mark the end of the last foot in a tone group need not be written. The first two tone groups in the example have a one syllable upbeat, two dactyls, and a final msf.[5] The next two have a two syllable upbeat, one dactyl, and a final msf. The last line departs from the pattern of the first two slightly, consisting of a two syllable upbeat, a dactyl, a trochee, and an msf.

[5] I use the terms *trochee* and *dactyl* from classical metrics, meaning two syllable (strong-weak) and three syllable (strong-weak-weak) foot respectively. Lacking an appropriate word for a one syllable foot, I use *msf,* for mono-syllabic foot (one strong syllable). I could have used *triseme* from classical metrics, but the implications of equal length with a trochee are incorrect, and besides the word is not familiar. Similarly I will use *qsf* (switching to Latin) for a four syllable foot (strong-weak-weak-weak), as *paeon primus* is hardly in common use.

One might well ask why we define the foot so that the salient syllable falls at the beginning (a descending foot), rather than the end (an *ascending* foot). This would indeed give a slightly simpler analysis of the above example in terms of iambs and anapests alone. There are a few reasons to prefer the descending analysis. There are limericks which have a simple descending analysis in terms of trochees and dactyls, but whose ascending analysis involves stranded non-salient syllables at the *end* of lines and msfs at the beginning. Consider the following limerick, first marked with descending feet, then with ascending, marked with backward slashes (\):

| / Bankers do / not work in / closets |
/ Most of them / say its be / cause its |
3a) */ Got to be / classy |*
/ Marbled and / brassy |
/ Else they would / get no de / posits |

| \ Bank \ ers do not \ work in clos \ ets |
\ Most \ of them say \ its because \ its |
3b) *\ Got \ to be class \ y |*
\ Mar \ bled and brass \ y |
\ Else \ they would get \ no depos \ its |

(3b) seems much more of an arbitrary, inappropriate division than 2 does, although admittedly I'm prejudiced. In particular, the initial msfs in (3b) seem much less reasonable than the final ones which follow from the descending analysis in (2).

Another argument follows from some facts presented by Bolinger [1965b] in service of another point, namely that in exaggeratedly rhythmic versions of e.g.

4a) */ Pa / made / John / tell / who / fired / those / guns*
4b) */ Pa / made the / man / tell / who / fired / those / guns*

the two sentences take the same length of time, with the time for the extra syllable in (4b) being provided by shortening *made*, with *man* taking about as long as *John*, suggesting that the dependency of non-salient syllables is to their left, which is to say, a descending foot analysis, as marked in the example. The fact that contractions adhere to their left provides the same sort of evidence for the choice of a descending foot. This example also demonstrates the basis for Halliday's reason (personal communication) for preferring the descending analysis, namely that an analysis of (4b) into descending feet is isochronous, while an ascending analysis is not (see below). The results of the transcription experiment as reported below in chapter 2, although not providing a direct comparison, at least demonstrate that the descending foot is not unnatural.

*1.1.2.1.5 **Isochronicity and stress timing***

English has been called an *isochronous stress timed* language, which is to assert that salient syllables occur at regular intervals, or, more particularly, that feet tend to be of equal duration, regardless of the number of syllables they are composed of. It is certainly possible to speak English this way, as in the exaggeratedly rhythmic (4a) and (4b) above, or in the stilted delivery of an amateur Shakespearean actor. But it is certainly not true of ordinary speech - as Morris Halle[6] has pointed out, the very fact that we can easily tell when someone is speaking to a strict rhythm demonstrates that one normally does not speak in that fashion. There is of course some truth to the isochronicity hypothesis, in that since salient syllables are frequently longer than their non-salient successors, it follows that a trochee will not be twice as long as an msf, nor a dactyl three times as long.[7] This subject is discussed at length, in connection with the presentation of relevant experimental data, in chapter 3.

1.1.2.2 Categorial/tonal

Within the structure imposed on the flow of speech by the organizational/temporal system, various distinctions are articulated. These distinctions result in a categorization of tone groups in terms of a vocabulary of *intonational words* and a categorization of feet within tone groups as *highlighted*, *nuclear*, or *ordinary*. These categorizations compose the categorial/tonal subsystem of prosody.

This part of prosody has received much more attention from linguists than the organizational/temporal, but it concerns me less here, and will accordingly be passed over rather lightly, except insofar as it interacts crucially with the organizational/temporal.

[6] cited by Mark Liberman, personal communication

[7] It is also possible that there is a definite difference here between American English and British English on this score. None of the claims made herein should be taken to apply to anything other than American English.

1.1.2.2.1 Envelope, boundary tones, kinetic tone: the intonational word

From the phenomenal perspective an *intonational word* is an element of a vocabulary of tonal patterns, just as an ordinary word is an element of a vocabulary of segmental patterns. From the functional perspective, both intonational and ordinary words convey meaning, broadly speaking. Taking a phonesthetic, ideophonic viewpoint, following Liberman and Ladd, I posit an intonational lexicon, whose elements are intonational words.

Given the assumption that the meaning, or at least the affect, of an utterance can be separated into that which comes from the words, syntax, and organizational/temporal structure on the one hand and the intonation on the other, then intonational words are what determine the space of possibilities for intonational form and content, and their association. For example (5a) and (5b) below are identical in all but the intonational word which is co-articulated with them:

5a) *John's coming.*
5b) *John's coming?*

In (5a) the pitch declines gently, and then turns down at the end, and the sentence is heard as an assertion, while in (5b) the pitch rises gently, and then turns up at the end, and the sentence is heard as a question.

From my perspective there are four aspects of an utterance which are prosodic and which are fundamentally tonal in nature: The shape of *kinetic tones*, the *tonal envelope* of the tone group, the sign and magnitude of the *boundary tones*, and the sign and magnitude of *tonal excursions*. An intonational word specifies all but the last. Thus a particular intonational word, such as the simple declarative contour which appears in (5a) above, specifies both an envelope, boundary tones, and a kinetic tone, on the one hand, and some sort of meaning schema on the other.[8]

Typically a tone group ends with a continuous smooth pitch change, usually over the last foot in the tone group. This tonal movement is called a *kinetic tone*. It usually moves down for statements and Wh-questions and up for yes-no questions in American English, although more

[8] Note that this approach differs from Liberman's in separating out tonal excursions, whereas e.g. Liberman and Sag's [1974] discussion of the *surprise-redundancy contour* includes the specification of what I would consider a tonal excursion as a necessary component of the contour.

complex 'tunes' are possible. In this context a downward pitch movement is called a *fall*, and an upward movement a *rise*. *Fall-rise* and *rise-fall* have the obvious meaning, and the kinetic tones they name also occur regularly in American English, although the specification of the meaning of the intonational words they enter into is by no means clear. Other kinetic tones are discussed in the literature: tripartite ones and the so-called 'defective' or 'level' kinetic tone, but they will not be of concern here. In addition to the obvious property of direction, kinetic tones may also differ in range, that is in how far they move up or down in their various parts. I take no position here on the question of whether there is a necessarily one to one relation between tone groups and kinetic tones: There must be at least one kinetic tone in a tone group, but it seems there may be two in some cases. See section 2.3 for discussion of this point in light of some experimental results.

In addition to the rather abrupt pitch movement of the kinetic tone, there is a gradual change in the floor and ceiling, as it were, of the pitch variations in an utterance. This is usually gently downwards, as in the declarative contour of (5a), but may be upwards in some cases, e.g. some versions of (5b), and is called the *tonal envelope* of the utterance. The envelope determines what counts as high or low pitch at any point, making such judgments relative to position within the tone group. For example, in a tone group whose intonational word specifies a falling envelope, a fundamental frequency which is perceived as a high pitch near the end of the tone group may in fact be lower in fundamental frequency than one perceived as a low pitch at the beginning of the same tone group. Pierrehumbert [1979, 1980] describes this phenomenon in detail with experimental evidence.

Finally, at both the beginning and the end of a tone group there may be an abrupt change in pitch, which can be conveniently described as oriented towards some target pitch at the tone group boundary, called a *boundary tone*, so that the first and/or last syllable of the tone group, whether salient or not, is noticeably high or low in pitch, with a suitable accommodation of the neighboring syllables.

As well as the strictly tonal aspects of the above three phenomena, there is a temporal aspect as well, namely their location within the tone group. This is the problem of *tune-text association*, that is, given the segmental specification of an utterance and an intonational word to be co-articulated with it, how do the pieces match up? This problem, and others dealing with the interactions between the categorial/tonal system and the organizational/temporal will be discussed in Chapter 4.

The questions of just what parts of the various aspects of the intonational word are distinctive, and what the nature of the 'meaning' of such 'words' is, have vexed linguists for some time, but beyond subscribing to the belief that there are such meanings, I have nothing to add to this debate here.

1.1.2.2.2 Tonal excursion and highlighting

At any point in an utterance, a syllable may be emphasized or highlighted by being displaced in pitch, either upwards or, more rarely, downwards, from the surrounding material. I refer to this as *tonal excursion*, and to its effect as *highlighting*. This is distinct from *salience*, as defined above, and is similar to what Bolinger means by *accent*. There is some suggestion that the direction of the tonal excursion is conditioned by the tonal envelope. Tonal excursions are usually upwards if the tonal envelope slopes downwards, and downwards if it slopes upwards, but this is not firmly established, and indeed as upward sloping envelopes are much rarer in American than in British English, downward excursions are also quite rare in American English.

In separating tonal excursion and highlighting from the intonational word, I stand in disagreement with much previous work on prosody. I have done this for two reasons. First, from the functional perspective, there does not seem to be any principled limit to the number of highlighted constituents - two is common (see e.g. [Jackendoff 1972]) and more perfectly acceptable. For example, the second line of the dialog given below has three, indicated by underlining and bold-face, and does not seem unreasonable:[9]

6) *Lee: "I hear that you gave Robin a bookcase."*
*Kim: "Nope, you got it wrong, **Mor**gan **bor**rowed a bookcase **from** Robin."*

Secondly, the relative importance of the two phenomenal properties location and tune is different for tonal excursions on the one hand and kinetic tones on the other. For tonal excursions, the principal semantic effect is determined by *where* they occur, and their sign (high or low) is of little importance, whereas for kinetic tones, it is whether they fall or rise which makes the greatest difference, and their location is much less important, usually being predictable from

[9] Note that not all highlighting is what is usually called *contrastive stress*, although this is one of the major functional roles of highlighting, as in (6). (3a) and (3b) in chapter 0 exemplify another, non-contrastive use.

the location of the end of the tone group. Another way of putting this distinction is that the principal choice made by the speaker for highlighting is what to highlight, while the principal choice for kinetic tone is which kinetic tone to use.

Thirdly, making this distinction allows a perspicuous answer to the vexatious question of whether the intonational lexicon distinguishes between e.g. *fall* and *high fall.* I think the right answer is *no*: The distinction in affect between the two intonations so described is better accounted for in terms of one intonational word, specifying a fall, together with its meaning, and the presence or absence of an upward tonal excursion, together with *its* contribution.

Finally, from the functional perspective the two clearly serve very different purposes - highlighting as against the specification of mood (in both the grammatical and the emotional senses of the word). Some distinction is called for on this basis alone, although whether the appropriate distinction from the phenomenal perspective is between kinetic versus static excursions is less clear - perhaps some slightly different approach would abandon that distinction for another, which would also resolve the question of how many kinetic tone there are in a tone group, namely a distinction between on the one hand the last tonal activity of any kind in a tone group, with a focus on its shape, which is a part of the intonational word, and on the other the location of all instances of tonal activity, with a focus on the unit picked out thereby, shape being secondary. In fact Pierrehumbert and Selkirk have recently made proposals along this line ([Pierrehumbert 1979, 1980], [Selkirk 1979, forthcoming]).

1.1.2.2.3 Notation

The notation for tonal phenomena which I will use is straightforward. It is distinguished from that for temporal phenomena by being placed above the relevant syllable, which will almost always be the first syllable in its foot. I use a backward slash (\ ) for a falling tune, a slash for a rising tune (/), and combinations for more complex tunes, e.g. V for fall-rise and Λ for rise-fall. Upwards tonal excursions are marked with an up-arrow (↑), and downwards excursions with a down-arrow (↓). Marks indicating excursions precede those for kinetic tones, so for instance ↑\ would be the notation for the high fall discussed above. As an example of all this, there follows a complete notation of one possible version of the limerick in (1) above, with both tonal and temporal features marked:

```
                                           \
       | I / knew a young / lady named / Bright |
                                           \
       Who / traveled much / faster than / light |
                                 /
6)     She de / parted one / day |
                             /
       In a / relative / way |
                             ↑           \
       And ar / rived the / preceding / night |
```

This concludes the introduction to phenomena and terms - Appendix 1 contains brief definitions of all the terms introduced here and some others, as well as attributions in some cases and pointers back into this section for the original definitions.

1.2 Questions and answers

Two fundamental questions arise from the approach laid out in the previous section, one concerning the phenomenal analysis proposed and one concerning the functional.

1.2.1 Getting at 'the fact of the matter': Transcription methodology

At the beginning of the previous section, I called attention to the essential circularity of phenomenal analysis in the prosodic arena. The history of the subject consists in large part of disagreements not just over how to analyze what is going on, but more fundamentally over what *is* going on. Most positions have been propounded in much the same style as I adopted above - this is how it is, take it or leave it. This has given a strong idiosyncratic flavor to the field, with almost as many positions as there are theorists. As Labov and Fanshell [1977] put it, "Our general point is that there is no general system for analyzing intonational contours that is generally accepted in linguistics in the same way that phonetic transcription is accepted." As far as I know, there have been only three cases where an effort has been made to stabilize and transmit an approach to notation and transcription in an effective way. The Trager and Smith [1951] system was developed within the context of American Structuralism, and viewed as an extension of the linguist's phonetic vocabulary. A significant effort was made to standardize notation, and to train linguists via a paradigm similar to that by which the IPA is taught. A sample is presented, the students transcribe, the teacher gives the 'correct' transcription, the students listen again and recognize their 'mistakes'. But certain aspects of the system, particularly the insistence on four distinct, absolute pitch levels, proved extremely difficult to transmit in this way. "The difficulty of teaching a pitch-level scheme to classes of students first led me to the feeling that one is dealing with contours instead of discrete pitch levels. Students often grasp the contours quickly, but have endless difficulty marking the levels." [Gunter 1974]. This in turn suggests one criterion at least for a system: It should be teachable. If it is not, that is strong, albeit circumstantial, evidence that the distinctions it makes are the wrong ones. Pike [1945] was another attempt at standardization, and included a phonograph record of examples in support of this goal, but the system was not a success. It was not in fact very different from the Trager-Smith system, and suffered from the same fundamental problems. Finally Trim [ms.] has prepared an excellent set of training materials, including a tape with many examples and exercises, but unfortunately it covers only the identification of kinetic tones, and does not deal with any other aspects of the problem. The comparison with Daniel Jones's recording of the cardinal vowels is apt, and indeed Trim's inventory does seem complete, and probably learnable.

This brings us to another criterion for a system: It should be learnable. This is really no different from the teachability requirement, but the focus is different. If a system is learnable, then it will give consistent results when applied to the same data on different occasions by the same subject, or by different, equivalently trained and equally facile subjects. Very little validation of theories in this way has ever been done, *pace* Lieberman [1965], which showed that well-trained and experienced linguists could not agree on tone level transcription *a la* Trager-Smith or Pike. Lea [1973, 1976] investigated the consistency across trials and subjects of marking syllables as either stressed, unstressed, or reduced. He achieved agreement as high as 95% for some subjects, but all his stimuli were read or recited, and there is some suggestion that the consistency drops significantly with natural, conversational data. He also includes no discussion of the training procedure used, if any. Li, Hughes, and Snow [1973] mention a similar experiment, with apparently somewhat less inter-subject consistency, perhaps around 70%, although they do not give the statistics in a form which allows a consistency measure to be determined.

All this suggests that a prosodic theory is not complete unless it includes some specification of a training procedure, and some indication that the procedure and the system are effective, that is, they produce consistent results across trials and subjects when applied to natural, unrehearsed speech. Note that it takes both consistency across trials for each subject, to demonstrate that the subject has learned something, and consistency across subjects, to show that more than a memory for previous responses is involved, that there is something in the stimulus which the subjects are responding to.

Chapter 2 below reports on an experiment designed to test a training method and transcription methodology based on the approach to prosodic phenomena presented in section 1.1 above. Of the phenomena defined above, subjects' ability to consistently mark feet, tone groups, kinetic tones and tonal excursion were tested.

1.2.2 Acoustic correlates of prosodic phenomena

For those prosodic phenomena my analysis of which is supported by subjective experimentation, the question of the acoustic basis, if any, for their discrimination remains. For instance, given a tape recording of a stretch of speech, together with a consensus judgment of its division into tone groups and feet, what acoustic characterization of these units emerges? What light, if any, is shed on the isochronicity hypothesis? Chapter 3 below examines some of these questions.

1.2.3 What's it all about, anyway?

An asymmetry in favor of the phenomenal should have been apparent throughout most of section 1.1 above. From the functional perspective, the roles of much of the acoustic structure discussed is less than clear. We are back to the issues discussed in Chapter 0 - the problem of how prosody functions to contextualize an utterance. There are two subparts to this problem, which I will call the *syntactic* and the *semantic*. Just as in the non-prosodic domain, these two are not totally separable, but there is some tactical advantage to considering them as if they were. The semantic problem is that of determining the meaning or affect of different prosodic forms. The obvious parallels to lexical semantics and word formation can be drawn: What are the contextual determinants of the choice of intonational word for a given utterance, and how can the contributions of various aspects of that word be distinguished and categorized.

The syntactic problem, which is the one with which I am more concerned here, is in some sense prior to the semantic one. It is concerned with determining what the possible prosodic forms for an utterance are. Some constraints along this line have already been mentioned: The composition of tone groups from an upbeat and a sequence of feet, the possible restriction to one kinetic tone per tone group. These are stated totally within the terms of the prosodic system, but there is another related class of questions concerning the constraints on the co-articulation of the prosodic and segmental aspects of an utterance. There appears to be some relation between syntactic boundaries and tone group boundaries - what is it exactly? What about tune-text association - what determines which syllable gets the kinetic tone? Does some interaction of contextual, semantic, syntactic, lexical, and phonological factors determine the division into feet, and if so, how?

My principal concern is with this last question. Most previous work on the syntactic and lexical parts of the question has been restricted to the assignment of abstract lexical stress on the one hand, or to questions of the syntactic conditioning of so-called 'sentence stress' on the other, which for me is not part of this issue, having to do rather with the location of the kinetic tone. About all that has been said in the past about *salience*, as opposed to *stress*, in conversation is that all other things being equal, the stressed syllable of 'content' words (or alternatively of 'open class' lexical items) will be salient (occur at the beginning of feet), and other syllables will not.

Even a cursory examination of conversational data show that, if taken literally, this hypothesis is false on all three of its predictions. Content words, particularly verbs, frequently are realized as non-salient. Closed class items such as prepositions, auxiliaries, and quantifiers frequently are realized as salient. And even the unstressed syllables of words of both classes are salient when they are the site of a tonal excursion as well.

On the other hand substantial progress has been made on issues of stress, both at the word and constituent level, most notably in [Liberman 1975] and [Liberman and Prince 1977]. The approach taken therein both to the phonological basis for, and relation nature of, lexical stress is a definite step forward, and together with a notion of an underlying rhythmic structure called the *metrical grid*, also provides considerable insight into the way in which the stress patterns of neighboring words may affect foot structure. In particular some cogent proposals have been made with respect to what has been called the rhythm rule or the 'thirteen men' rule. This concerns itself with the relation of e.g. *artificial*, which has stress on its third syllable, to the frequently observed foot boundary pattern of */ artificial in/telligence*, where the foot boundary in *artificial* has been retracted from its 'expected' position in front of the stressed third syllable.

The time seems ripe for a more careful and principled approach to the salience problem than has appeared heretofore. Starting from the improved account of lexical stress, and its accompanying formulation of the rhythm rule given in [Liberman and Prince 1977], Chapter 4 attempts to develop a processing model of the relevant aspects of speech production which specifies those contextual, semantic, syntactic, and phonological factors which may have an affect on foot structure, and shows how they determine the division into feet of an utterance. The model is based primarily on the consensus foot boundary assignment which emerged from the experiment described in Chapter 2, and the measurements on those feet reported in Chapter 3.

Chapter 2. The transcription experiment

In which the transcription experiment alluded to in the preceding chapters is described in detail and its results presented and discussed.

2.1 Method

2.1.1 Subjects

The principal subjects for the experiment were four graduate students in the linguistics department at the University of California at Berkeley, three female and one male, all native speakers of English. All had some prior interest in and experience with natural discourse and intonation, but none had any previous experience with close transcription of prosody. Two of the subjects (D and C) were paid for their time; the other two volunteered without pay out of interest in the issues involved. I also performed most aspects of the experiment, although not in exactly the same form as the others.

2.1.2 Materials

The basic material for the experiment consisted of two segments of natural monologue, the *training segment*, 33 seconds long, and the *data segment*, 130 seconds long, both of which are contained on the accompanying soundsheet.[1] Both were excerpted from a single recording obtained in the following way. The speaker is a 36 year old white male, native speaker of English with native speaker parents, born in Mobile, Alabama and raised mostly in Texas, without significant regional accent. He is a research scientist in computer science, a colleague who participated as a favor. The original recording was made using an AKG D110 200 ohm lavaliere microphone hung around the speaker's neck and a Sony TC 854-4 open reel tape recorder running at 15 ips. It was made in the speaker's office, with the door closed, and is of very high quality, with very little background noise present and the speaker's voice clear and distinct. After arranging the equipment, starting the recording, and eliciting the background data given above, I explained to the speaker that I was interested in various (unspecified) aspects of monologue, and was going to ask him a few questions designed to provoke an extended response

[1] As the soundsheet is not as hard as a non-flexible disc, repeated playing will wear it out. I recommend copying the texts to a cassette if you anticipate repeated use of the material

from him without further intervention on my part. The first question I asked was "Did you have any sort of particular family rituals associated with Christmas?" After some negotiation about the topic, the speaker began to respond. After several false starts, he said "I don't know there was any great ritual but I'll just tell you about christmas um". The data segment then follows for the next 130 seconds, and, after a few clauses, the training segment brings the monologue to a close (see below for the contents of the segments). During the course of the recording the speaker was facing me, and the tape recorder was out of sight. Although I was careful to make no verbal response to his story, I played the role of an interested listener, responding with nods, smiles, and other non-verbal regulators when appropriate. As near as I could tell, the speaker was relaxed, relatively unaware of and unaffected by the microphone and tape recorder, neither of which he could see, and to all intents and purposes engaged in an interesting description of his childhood to an interested hearer, for its own sake. All this is to say that the results qualify as ordinary, natural speech, and all those who have heard the tape agree that it is.

I copied the training and data segments onto cassettes for use in the experiment, with a small but not particularly noticeable loss in fidelity. I also transcribed the contents of the two segments on a syllable by syllable basis, including false starts and filled pauses, but using English orthography.

Using the syllable level transcription as a worksheet, on four separate occasions I transcribed the prosodic properties of the training segment which were of concern, namely the division into feet, the division into tone groups, the kinetic tones, and the tonal excursions. Marking foot boundaries was done as one task on one worksheet, but the other three properties were done together on a second worksheet. Comparing the results of those trials and checking again with the tape, I arrived at consensus notations for the training segment, which are reproduced below, in the form in which they were subsequently given to the subjects. Areas of residual uncertainty in transcription are indicated by underlining.

Sample a

This example demonstrates the foot boundary transcription style with areas of low confidence underlined.

I / guess we / of-ten / went there I / guess we / of-ten / went there / on uh / thanks-gi-ving / ac-tu-al-ly uh I sup/pose but / I re-mem-ber these / big tra/di-tion-al / thanks-gi-ving / din-ners with / tur-key and / dres-sing which / I ne-ver / liked and and uh / cran-ber-ry uh / sau-ces and / lots and / lots of / pies and / lots and / lots of / can-dies and / she / made / spe-cial-ly her / can-dies I re/mem-ber and / le-mon me/ringue / pies and in/cre-di-ble / stuff / good good / stuff um so / that was / al-ways ex/ci-ting

Figure 1. Transcription of foot boundaries of training segment, as given to subjects

Sample b

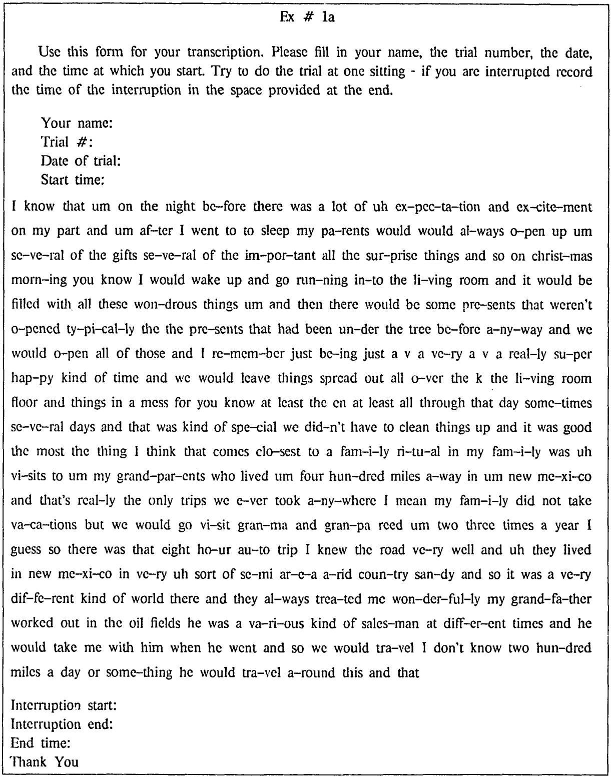
This example demonstrates the tone group boundary, etc. transcription style with areas of low confidence underlined.

```
                        \         *                                ↑\                 /
I guess we of-ten went there | I guess we of-ten went there on uh thanks-gi-ving | ac-tu-al-ly |
      \                                                                      \               \
uh I sup-pose | but I re-mem-ber these big tra-di-tion-al thanks-gi-ving din-ners | with tur-key
       \                          \                                 \
| and dres-sing | which I ne-ver liked | and and uh cran-ber-ry uh sau-ces | and lots and lots
    \                               \                                 ↑\             /
of pies | and lots and lots of can-dies | and she made spe-cial-ly her can-dies I re-mem-ber |
                      \             ↑         \                 \             ↑
and le-mon me-ringue pies | and in-cre-di-ble stuff | good good stuff | um so that was al-ways
   /
ex-ci-ting |
```

*This word and those following are definitely elevated in pitch, but it seems to be more of a register or envelope phenomenon than the kind of pitch intrusion we mark with '↑'. You can try to mark this kind of thing in with subjective contours stretching over some distance, if appropriate.

Figure 2. Transcription of tone group boundaries, kinetic tones, and tonal excursions of training segment, as given to subjects

On the next page, the worksheet for the data segment is shown. The heading on the sheet varied to suit the particular subtask it was to be used for. There is a small error in this worksheet in that the word *only*, in the phrase *the only trips we ever took anywhere*, about two thirds of the way through, does not have a syllable break included - none of the subjects noticed this, so I do not think it affected the results.



Ex # 1a

Use this form for your transcription. Please fill in your name, the trial number, the date, and the time at which you start. Try to do the trial at one sitting - if you are interrupted record the time of the interruption in the space provided at the end.

Your name:
Trial #:
Date of trial:
Start time:

I know that um on the night be–fore there was a lot of uh ex–pec–ta–tion and ex–cite–ment on my part and um af–ter I went to to sleep my pa–rents would would al–ways o–pen up um se–ve–ral of the gifts se–ve–ral of the im–por–tant all the sur–prise things and so on christ–mas morn–ing you know I would wake up and go run–ning in–to the li–ving room and it would be filled with all these won–drous things um and then there would be some pre–sents that weren't o–pened ty–pi–cal–ly the the pre–sents that had been un–der the tree be–fore a–ny–way and we would o–pen all of those and I re–mem–ber just be–ing just a v a ve–ry a v a real–ly su–per hap–py kind of time and we would leave things spread out all o–ver the k the li–ving room floor and things in a mess for you know at least the en at least all through that day some–times se–ve–ral days and that was kind of spe–cial we did–n't have to clean things up and it was good the most the thing I think that comes clo–sest to a fam–i–ly ri–tu–al in my fam–i–ly was uh vi–sits to um my grand–par–ents who lived um four hun–dred miles a–way in um new me–xi–co and that's real–ly the only trips we e–ver took a–ny–where I mean my fam–i–ly did not take va–ca–tions but we would go vi–sit gran–ma and gran–pa reed um two three times a year I guess so there was that eight ho–ur au–to trip I knew the road ve–ry well and uh they lived in new me–xi–co in ve–ry uh sort of se–mi ar–e–a a–rid coun–try san–dy and so it was a ve–ry dif–fe–rent kind of world there and they al–ways trea–ted me won–der–ful–ly my grand–fa–ther worked out in the oil fields he was a va–ri–ous kind of sales–man at diff–er–ent times and he would take me with him when he went and so we would tra–vel I don't know two hun–dred miles a day or some–thing he would tra–vel a–round this and that

Interruption start:
Interruption end:
End time:
Thank You

Figure 3. Transcription worksheet for the data segment

2.1.3 Procedure

Each of the four subjects was given a packet of materials consisting of a copy of the training and data segments on a cassette, the sample transcriptions of the training segment shown above as Figures 1 and 2, twelve worksheets like the one shown in Figure 3, and two pages of instructions. In accord with those instructions, which appear on the following pages, there were four worksheets with the heading 'Ex # 1a', four with the heading 'Ex # 1b', and one each with headings 'Ex # 1a with confidence rating', 'Ex # 1b with confidence rating', 'Ex # 1a Summary', and 'Ex # 1b Summary'.

This is a two part experiment in transcribing prosody from live data - one part for accent and one for intonation. The basic procedure is the same for each of the two parts, and I will begin by describing it. Then the details of each task will be described.

Basic Procedure

1. Read the directions for the foot boundary task, #1a (below).

2. Listen to the sample and look at Sample a. Stop and start as needed, listen several times to get familiar with the system. The cassettes are CrO2, stereo, but you can play them on anything. I recommend using earphones if you can, it cuts down on distractions.

3. Listen to the text itself once, thinking about the task at hand.

4. Transcribe the text once on one of the forms provided. Stop and start as needed, back up if you wish, but try to keep your momentum - if you are stuck over a difficult spot, make your best guess and press on. Don't spend more than about 15 minutes on the foot boundary task and 10 minutes on the tone task, but on the other hand don't watch the clock.

5. Repeat steps 1 - 4 for the tone task, #1b.

6. Do something else. You should break for at least fifteen minutes, and probably not more than a day, or you will lose context.

7. Repeat steps 4 - 6 three times, transcribing feet and then tone and then taking a break. Do not look at your prior transcriptions, but do not actively ignore them either. That is, learn from experience, but do not refer to it overtly.

8. Do the transcriptions a fifth time, but this time also underline those parts of the text where you had difficulty making up your mind, that is where your confidence in the transcription is low. Examples of this are given in the samples. Go back and listen to the sample if you are unclear on this. Do the underlining carefully, so the domain of uncertainty is clear.

9. Do the tasks one last time, but this time look at all your previous transcriptions as you go and try to 'get it right'. That is, pay particular attention to places where your prior transcriptions do not agree and try to make up your mind. Again, underline areas where, despite your best efforts, the 'right' answer just isn't clear. Do not 'correct' your previous transcriptions.

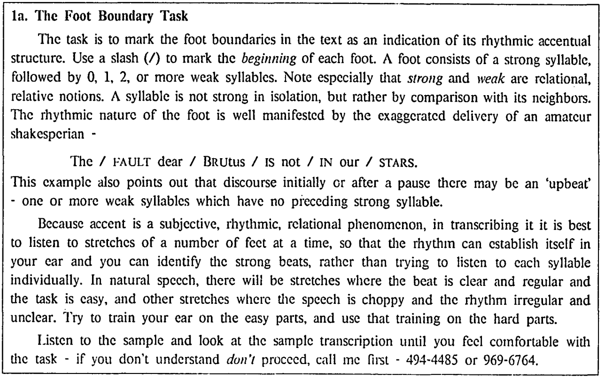
10. Buy yourself a beer at my expense.

11. Append any comments about the instructions, the task, etc. which you think would help me smooth this out.

Summary

Read the directions
Look at the samples
Do both transcriptions 4 times, with breaks in between
Do the transcriptions again, marking areas of low confidence
Do the summary transcriptions, referring to the previous ones, again marking low confidence areas

Figure 4a. First part of instructions



1a. The Foot Boundary Task

The task is to mark the foot boundaries in the text as an indication of its rhythmic accentual structure. Use a slash (/) to mark the *beginning* of each foot. A foot consists of a strong syllable, followed by 0, 1, 2, or more weak syllables. Note especially that *strong* and *weak* are relational, relative notions. A syllable is not strong in isolation, but rather by comparison with its neighbors. The rhythmic nature of the foot is well manifested by the exaggerated delivery of an amateur shakesperian -

The / FAULT dear / BRUtus / IS not / IN our / STARS.

This example also points out that discourse initially or after a pause there may be an 'upbeat' - one or more weak syllables which have no preceding strong syllable.

Because accent is a subjective, rhythmic, relational phenomenon, in transcribing it it is best to listen to stretches of a number of feet at a time, so that the rhythm can establish itself in your ear and you can identify the strong beats, rather than trying to listen to each syllable individually. In natural speech, there will be stretches where the beat is clear and regular and the task is easy, and other stretches where the speech is choppy and the rhythm irregular and unclear. Try to train your ear on the easy parts, and use that training on the hard parts.

Listen to the sample and look at the sample transcription until you feel comfortable with the task - if you don't understand *don't* proceed, call me first - 494-4485 or 969-6764.

Figure 4b. Second part of instructions

1b. The Tone Task

This one is a bit more complex as there are three somewhat different things going on - tone group boundary marking, contour marking, and tonal excursion marking.

The intonational structure of a text is subjectively quite clear, although difficult to justify. The words of a text group together in intonational units or phrases, which often, although not always, correspond with syntactic constituents. There are often pauses at tone group boundaries, and there is usually a tonal contour or 'kinetic' (moving) tone. But both pauses and contours may occur elsewhere as well, so there appearance is not diagnostic. Tone group boundaries are closely related to the mysterious classical notion of 'juncture' - you know it's there even though you may not be able to say why. Use a vertical bar (|) to separate tone groups in the transcription. Listen to the sample and look at sample b and try to get a sense of what it is that the vertical bars are marking.

Use slash (/) and back-slash (\) to mark local contours - that is, contours which occur within a single syllable or across a disyllable (occasionally over a longer stretch). You may need to compound these to notate e.g. a rise-fall (/\). There is no prior constraint on what contours may occur. The only contours which occur in the sample are the familiar declarative fall and suspensive rise. Note also that contours are not restricted to tone group boundaries or to occur one per tone group.

Use up-arrow (↑) and down-arrow (↓) to mark tonal excursions - syllables which are tonally prominent either upwards or downwards but not part of contours. This is really a subjective category - it rests on a subjective feeling that the ordinary appearance of contours, especially at the end of tone groups, is distinct from the use of tonal prominence to in some way emphasize a constituent. Again, the sample should make both the clear cases and the problems evident. Note that these last two systems interact, so that a high or emphatic fall is notated (↑\) as opposed to an ordinary fall (\).

Again, work with the sample until you feel comfortable and call me at need.

Figure 4c. Third part of instructions

All four subjects did in fact call at some point during the preliminary stages, all with more or less the same problem - there were a few places where they disagreed with the sample transcriptions. My response was the same in all cases, namely that if they understood the system well enough to know that they disagreed in a particular specific place, then the training exercise had in fact been successful, and they could proceed to the data segment. This was not premeditated, in that the errors were not inserted on purpose, but turned out well, as it established both for me and the subjects that the training procedure did have some definite effect.[2]

Other than that subjects seemed to have no difficulty in understanding or following the directions, although one subject (D) did all but the first pair of transcriptions in one eighty minute stretch without a break, thus not following instruction 6. More on this below.

I myself also did some form of the experiment, in part as I was developing its exact contents. In the end I did four independent trials, and one summary trial some time later, for both subtasks. I include my results in the statistics not because they were derived from the same strict experimental conditions, but rather because they give some measure of the success of the training procedure, in so far as to the extent that the other subjects agree with me, I have succeeded in communicating what underlies my own perceptions via the training procedure.

On receiving the worksheets back from a subject, I entered the results into a computer system as a preliminary to the statistical analyses which are presented below. Appendix C describes the interactive computer system I developed to aid in the task of entering the results, which was a task of considerable magnitude, as there were an average of 150 foot boundaries marked on each of 29 worksheets[3] together with an average of 80 tone group boundaries, kinetic tones, and tonal excursions marked on the other 29 worksheets, for a total of approximately 6700 data points. A pair of representative worksheets are reproduced as I received them in Appendix B.

[2] The disagreements were mostly about the presence of foot boundaries before the second and fourth instances of *lots* in the phrase *lots and lots of pies and lots and lots of candies*, which I had marked in the sample but which the subjects, and I in retrospect, did not think were there, and the choice of kinetic tone in a few places.

[3] Six each from four subjects and five from me.

2.2 Results and discussion

This section presents the results of the transcription experiment. Large numbers of statistics are presented, at a level of detail not often found in 'non-hyphenated' linguistics. I feel strongly that in this case this detail is appropriate and indeed necessary. As discussed above, one of the major difficulties besetting the study of prosody is an inability to compare the results of different approaches. My goal in presenting so much statistical material, and in being so explicit as to the exact basis of these statistics, is to make such comparisons possible in the future, both by providing a set of statistics which others can compare their results to, and by defining a set of procedures whereby others can analyze their data so as to yield comparable statistics of their own with which to make the comparison.

All the statistical calculations on the results of the experiment were performed with the help of the Interactive Data-analysis Language system, which is described briefly in Appendix C.

The data for each of the four sub-tasks was represented for analysis in a relatively uniform way. The data segment consists of 450 syllables, which gives 451 possible sites for boundary markings, and 450 possible sites for kinetic tone or tonal excursion markings. The results for each trial of the two boundary tasks were represented by a list of 451 zeros or ones, with a one indicating that the corresponding syllable had a boundary marked before it, and a zero that none was marked. The results for each trial of the location and identification tasks were similarly encoded by a list of 450 entries, each either zero, indicating no mark on the corresponding syllable, or a number encoding the mark made.

The basic measure of concern for all the tasks is the percentage agreement between pairs of trials. The agreement matrix between trials for the same subject gives some indication of whether the subject is learning as he goes along, or is simply marking at random. An increase in the agreement between one trial and the next as the experiment proceeds is a good sign in this regard, as is increasing agreement with the last or summary trial. Thus in the agreement matrices which follow the major points of interest are the uppermost diagonal, which give the successive agreements between one trial and the next, and the bottom row, which gives the agreement of each trial with the last.

Good results here are not in and of themselves sufficient however. The subject may have simply been getting better at remembering what he did on the previous trial, after having marked essentially at random the first time. To establish that something in the data itself is reflected in the results, a good agreement between subjects is also required. I have given only the agreements across subjects for the final summary trial of each task, assuming that that represents most accurately each subject's opinion.

There is some question in my mind as to what the appropriate null hypothesis is with respect to which to evaluate the results. I started out considering all 451 sites, but I think this led to overly optimistic statistics. For instance, for the foot boundary marking task, over half the sites were never marked by any subject on any trial. Allowing this trivial agreement of non-marking at 50% of the sites to artificially inflate the agreement between trials clearly obscures what is going on. This effect is even more serious for tone group boundaries, where over 80% of the sites were never marked. Although it is clear that concern should be restricted to a subset of the sites, what the subset should be is not clear. I have noted above that in this arena there is no 'fact of the matter', no a-priori basis for choosing the subset of relevance. Given this indeterminacy, I have gone to the pessimistic extreme, that is, the one which yields the lowest possible percentages. For each subject, I consider only the subset of sites which that subject ever marked for the task of concern.

It follows that the procedure for calculating the agreement matrix for a given subject for e.g. the foot boundary task is as follows. First identify those sites out of the 451 which the subject marked on at least one trial. Call these the *sites at issue* for that subject. The percentage agreement between two trials for that subject is then the percentage of the sites at issue for that subject for which the two trials are in agreement, either both marking a boundary or both not marking. The percentage agreement for all pairs of trials for a given subject make up the agreement matrix.[4]

Because of the way these matrices are computed, the null hypothesis is not immediately obvious. The question is, what kind of percentage agreement would you expect if the subject were marking at random. We can get a good idea of the answer if we consider the average number of marks made as giving the probability that the subject will mark any given site. By working through a simplified example, we can see what the answer is in the general case.

[4] The extreme conservatism of this approach should be born in mind when comparing these results to others published elsewhere. I hope others will include measures computed in the same manner in future work, to allow sensible comparison. It is unfortunate, for example, that it is impossible to sensibly compare my results with those presented in [Lea 1973] because the statistics he presents are not sufficiently detailed with respect to these issues.

Suppose there are only 36 sites, and the subject marks an average of 12. This gives a probability of 12/36, or one third, that a given site will be marked on a given trial. For two independent trials, of the 12 sites marked in the first trial, 1/3 will be marked in the second, giving $12 \cdot 1/3$, or 4, as the number of sites which will be marked on both trials. This gives us $12 - 4$, or 8, marked on the first but not the second trial, and similarly 8 marked on the second but not the third, for a total of $4 + 8 + 8$, or 20, sites marked at least once on one of the trials. Thus the number of agreements is 4, the number of sites at issue is 20, and the agreement under the assumption of random marking is 4/20 or 20%. For the general case if we have n sites, with an average of m marks per trial, then we have $m \cdot m/n$ agreements out of $m \cdot m/n + 2(m - m \cdot m/n) = (2mn - m^2)/n$ sites at issue, for $m/(2n - m)$ percent agreement as our null hypothesis. This figure will be given with all agreement matrices presented below.

The details of these agreement matrices and the statistics particular to each task are discussed in the appropriate section below.

2.2.1 Foot boundary assignment

The agreement matrices for the marking of foot boundaries are given on the next page. Note that I did one less trial than the other subjects, and that the last trial (#5 for me, #6 for the others) is the summary trial.

On the page after that are graphs comparing the main diagonals and bottom rows of the agreement matrices.

Subject H 188 sites at issue, null hypothesis .232

	Trial			
Trial	1	2	3	4
2	.910			
3	.899	.883		
4	.915	.920	.888	
5	.915	.899	.888	.915

Mean = .904
Variance = .000

Subject A 203 sites at issue, null hypothesis .219

	Trial				
Trial	1	2	3	4	5
2	.764				
3	.793	.793			
4	.749	.798	.857		
5	.754	.783	.842	.867	
6	.793	.823	.872	.887	.901

Mean = .815
Variance = .002

Subject B 212 sites at issue, null hypcthesis .219

	Trial				
Trial	1	2	3	4	5
2	.651				
3	.684	.825			
4	.627	.741	.821		
5	.608	.741	.802	.915	
6	.627	.769	.830	.896	.943

Mean = .752
Variance = .011

Subject C 211 sites at issue, null hypothesis .225

	Trial				
Trial	1	2	3	4	5
2	.739				
3	.806	.829			
4	.749	.820	.839		
5	.791	.815	.872	.844	
6	.787	.801	.896	.877	.882

Mean = .822
Variance = .002

Subject D 149 sites at issue, null hypothesis .118

	Trial				
Trial	1	2	3	4	5
2	.685				
3	.671	.691			
4	.698	.705	.758		
5	.678	.617	.685	.805	
6	.691	.644	.711	.752	.758

Mean = .696
Variance = .002

Table 1. Percentage agreement between trials of foot boundary placement, for each subject, relative to the sites at issue for that subject.

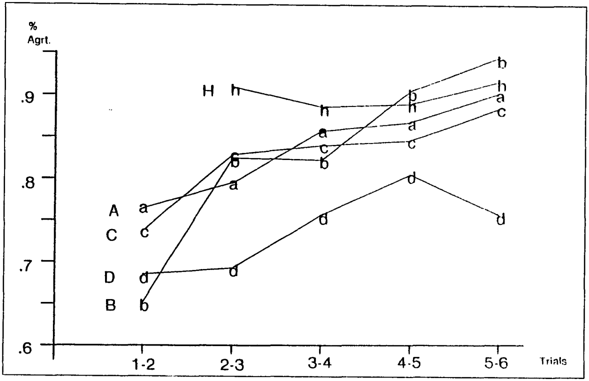


Figure 1a. Learning curves for Table 1, main diagonal (one trial vs. the next).

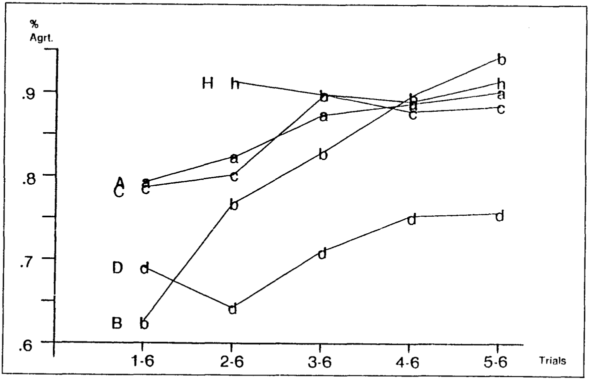


Figure 1b. Learning curves for Table 1, bottom row (each trial vs. summary).

The agreement matrices in Table 1, although not overwhelming, are quite good. The contrast between the figures for me and the others is interesting in several respects. I was clearly not improving as I went along - there is no learning manifested in the diagonals or bottom row. This is corroborated by the low variance. For the others, however, learning is apparent, although at different rates. This, and other comparative observations, are brought out most clearly in the graphs in Figures 1a and 1b.

Subjects A and C are quite similar, with steady learning, and a final cell near my typical values, suggesting that the six trials brought them close to a plateau, although further testing would be necessary to confirm that.

Subject B did less well. S/he starts off quite badly, but over the last three trials improves markedly, suggesting some early confusion which subsequently disappeared.

D did not do well at all. S/he shows erratic progress on the main diagonal and bottom row, and his/her best showing (80.5%) is well below that of the others. His/her overall mean is also by far the lowest. S/he also marked substantially fewer boundaries than the others, especially towards the end of each trial. The timings given on the worksheets are less than those given by the others, and no rest was taken between trials, all together suggesting that perhaps s/he did not pay quite as much attention as did the others.

To give some indication of the effect of the conservative approach I have taken to the statistics, the matrix below gives the agreement figures for subject A if all 451 sites are considered. Comparing this to the corresponding figures in Table 1, the inflation in the results is apparent, ranging from 14 percentage points at the low end to 6 percentage points at the high end. This differential aspect of the inflation is particularly pernicious, as it compresses the range of the data (from 15 percentage points to 7 percentage points separating the highest from the lowest) - note that the variance has dropped below .0005.

Subject A	all 451 sites at issue, null hypothesis .539				
	Trial				
Trial	1	2	3	4	5
2	.894				
3	.907	.907			
4	.887	.909	.936		
5	.889	.902	.929	.940	
6	.907	.920	.942	.949	.956
					Mean = .918
					Variance = .000

Table 2. Percentage agreement between trials of foot boundary placement, for subject A, relative to all 451 sites.

The next table shows the number of foot boundaries marked by each subject in each trial. Note that as mentioned above, subject D marked significantly fewer boundaries than anyone else ($p<.001$). Note that subject B's first trial differs significantly from the ones which follow ($p<.001$).

	Trial							
Subject	1	2	3	4	5	6	mean	var
H	167	170	170	175	167		169.8	10.7
A	149	149	161	168	173	173	162.2	123.4
B	122	164	161	177	175	175	162.3	433.5
C	181	162	166	158	163	164	165.7	63.5
D	84	69	93	97	114	116	95.5	321.1

Table 3. Total number of foot boundaries marked.

The next three tables give a different perspective on the consistency of the subjects. The first gives the number of sites which received a given number of boundary marks over all the trials for each subject. Those sites marked six out of six times[5] are a good sign, while those marked three out of six are indications of indecision.

[5] Or five out of five for me

	Number of times foot boundary marked						
Subject	0	1	2	3	4	5	6
H	263	12	3	10	14	149	
A	248	23	9	13	13	29	116
B	239	23	9	23	19	40	98
C	240	25	19	8	12	23	124
D	302	33	13	15	18	23	47

Table 4a. Number of sites receiving a given number of marks over all trials

The next two tables give percentage versions of the same results, considering only the sites at issue for a given subject. Table 4b collapses one mark out of six and five marks out of six in the column labelled 1 out, being those sites which were one shy of unanimity. Simiiarly, two out of six and four out of six are collapsed as 2 out. Three and three is Worst, and six out of six is Perfect.

	Consistency of foot boundary marking			
Subject	Perfect	1 out	2 out	Worst
H	.793	.138		.069
A	.571	.256	.108	.064
B	.462	.297	.132	.108
C	.588	.227	.147	.038
D	.315	.376	.208	.101

Table 4b. Percentage of sites at issue receiving a given number of inconsistent marks over all trials

Table 4c collapses even further, combining the unanimous and one out cases as against the worst and three out cases. The perspective provided by these three tables confirms the distinction made on the basis of the agreement matrices: Subjects A and C are noticeably surer of their marking than subject B, and subject D is even less consistent, being reasonably sure of less than 70% of his/her markings.

	Consistency of foot boundary marking	
Subject	1 or less	2 or more
H	.931	.069
A	.828	.172
B	.759	.241
C	.815	.185
D	.691	.309

Table 4c. Collapsed percentage of sites at issue receiving a given number of inconsistent marks over all trials

Having gotten some idea of the agreement across trials for each subject independently, the next tables address the question of agreement across subjects. Tables 5a and 5b are the agreement matrices for the last or summary trial of each subject. They differ in terms of what the sites at issue were. Table 5a is based on the sites at issue over all 29 trials of the foot boundary marking task. If any subject marked a site on any trial, then it is one of the sites at issue for this table. Table 5b takes a more conservative approach, considering only those sites marked in one of the five summary trials themselves.

258 sites at issue, null hypothesis .201

	Subject			
Subject	H	A	B	C
A	.853			
B	.736	.729		
C	.849	.857	.717	
D	.756	.725	.624	.736

Mean = .758
Variance = .006

Table 5a. Percentage agreement of summary foot boundary placements across subjects, sites at issue over all trials

223 sites at issue, null hypothesis .214

Subject \ Subject	H	A	B	C
A	.830			
B	.695	.686		
C	.825	.834	.673	
D	.717	.682	.565	.695

Mean = .720
Variance = .008

Table 5b. Percentage agreement of summary foot boundary placements across subjects,sites at issue over summary trials

These across subject agreement figures support the distinction drawn on the basis of the across trials figures given previously. Subjects H, A, and C agree substantially better with each other than B or D do with anyone. Between B and D the agreement is the worst of all.

The next table shows the agreement figures if we consider only H, A, and C.

195 sites at issue, null hypothesis .229

Subject \ Subject	H	A
A	.805	
C	.800	.810

Mean = .805
Variance = .000

Table 6. Percentage agreement of summary foot boundary placements across summary trials of H, A, and C, sites at issue over just those trials

We see from this that these three are indeed in substantial agreement.

The next two tables give the absolute and percentage consistency figures for the summary trials. Table 7a is for all five summary trials, Table 7b just for the three best subjects' summary trials.

	Number of times foot boundary marked					
	0	1	2	3	4	5
	228	39	25	19	51	89
%		17.5	11.2	8.5	22.9	39.9

Table 7a. Number of sites receiving a given number of marks over the five summary trials.

	Number of times foot boundary marked			
	0	1	2	3
	256	24	33	138
%		12.3	16.9	70.8

Table 7b. Number of sites receiving a given number of marks over the summary trials of subjects H, A, and C.

Once again we see evidence of the close association of H, A, and C: They mark unanimously on over 70% of the sites at issue for them, while including B and D reduces the figure to less than 40%. This is not because B and D agree on a different set of boundaries from the other three: They mark unanimously on only 50% of the sites at issue for the two of them.

The last question to be answered is: Given all these wonderful numbers, where are the foot boundaries in this text? I think I am justified in saying that the best answer to that lies pretty close to the 171 sites marked by two or more of subjects H, A, and C on their summary trials. The next page shows graphically the union of these three trials. The sites with a simple slash marked were agreed on by all three subjects. Sites marked by only two subjects are indicated by a slash with the initials of those marking above and below it, e.g. $^{a}/_{c}$. Sites marked by only one subject are indicated by a slash with the initial of that subject above it, e.g. $^{a}/$.

ᵃ⁄c I / know that um on the / night be/fore there was a / lot of uh ex-pec/ta-tion and ex/cite-ment on ᵃ⁄c my / part and um / af-ter / I went to to / sleep my / pa-rents would would / al-ways / o-pen / up um / se-ve-ral of the / gifts / se-ve-ral of the im/por-tant / all the sur/prise ʰ⁄ things and / so on / christ-mas / morn-ing you know ʰᵃ⁄c I would ᵃ⁄ wake / up and go / run-ning in-to the / li-ving room and / it would be / filled with / all these / won-drous / things um ʰ⁄ and ʰ⁄ₐ then ᵃ⁄ there would ʰ⁄ be / some ʰ⁄ pre-sents that / weren't ʰᵃ⁄c o-pened / ty-pi-cal-ly the the / pre-sents that ᵃ⁄ had ʰᵃ⁄c been ᵃ⁄c un-der the / tree beʰᵃ⁄cfore / a-ny-way and ᵃ⁄c we would / o-pen / all of / those and / I reᵃ⁄mem-ber just ʰᵃ⁄c be-ing ᵃ⁄ just a ʰ⁄ₐ v a / ve-ry a v a / real-ly / su-per / hap-py / kind of / time and ᵃ⁄c we would / leave ʰ⁄ₐ things ʰ⁄ₐ spread ᵃ⁄ out all ᵃ⁄c o-ver the k the / li-ving room / floor and / things in a / mess ʰ⁄ for ᵃ⁄c you know at / least the en at / least all / through ᵃ⁄ that / day / some-times / se-ve-ral ʰ⁄ₐ days and / that was / kind of / spe-cial we / did-n't ᵃ⁄ have to / clean ᵃ⁄c things / up and ᵃ⁄ it was / good the / most the / thing I / think that / comes / clo-sest to a / fam-i-ly / ri-tu-al ᵃ⁄ in ʰᵃ⁄c my / fam-i-ly ʰ⁄ₐ was uh / vi-sits to um my / grand-par-ents who / lived um / four hun-dred / miles a/way in um new / me-xi-co and / that's / real-ly the / only / trips we / e-ver ᵃ⁄ took / a-ny-where I mean my / fam-i-ly did / not ᵃ⁄ take va/ca-tions but ᵃ⁄c we would / go ʰ⁄ₐ vi-sit / gran-ma and / gran-pa / reed um / two ʰᵃ⁄c three ʰ⁄ₐ times a / year I ʰᵃ⁄c guess ʰ⁄ₐ so ᵃ⁄c there was ᵃ⁄ that / eight ho-ur / au-to ʰ⁄ trip I / knew the / road / ve-ry / well and uh ʰ⁄ₐ they / lived in new / me-xi-co in / ve-ry uh sort of / se-mi / ar-e-a / a-rid / coun-try / san-dy ʰ⁄ and ᵃ⁄ so it was a / ve-ry / dif-fe-rent / kind of / world ʰ⁄ₐ there and / they / al-ways / trea-ted ᵃ⁄ me / won-der-ful-ly ᵃ⁄ my / grand-fa-ther ʰ⁄ₐ worked ᵃ⁄c out in the / oil fields ᵃ⁄c he was a / va-ri-ous ʰ⁄ₐ kind of / sales-man at / diff-er-ent / times and / he would / take me / with him when he ʰ⁄ went and / so / we would / tra-vel / I don't know / two hun-dred / miles a / day or / some-thing he would / tra-vel a/round / this and / that

Figure 2. The union of foot boundary marking of subjects H, A, and C.

The next page shows as close to the truth as I think it's possible to get - the foot boundaries are those marked by two or more of H, A, and C, with three exceptions. I have included my marks before *and* at the beginning of the sixth line, before *for* at the end of the tenth line, and before *and* at the beginning of the fifth line from the bottom. All three of these markings are before long, drawn out, space filling words, which are separated from the words around them. They are in a sense defective feet, but for consistency I have marked them. The instructions did not discuss such cases, and nothing similar was contained in the training segment, so the lack of agreement here is not surprising.

The boundaries shown below in Figure 3 will be used in section 2.2.4 below as defining the feet of the text, as a complete partition of the text is required. In chapter 3, however, a more conservative set, namely that given by the unanimously marked feet, is used.

General discussion of the results of the foot boundary marking task will be found in section 2.3 below.

/ I / know that um on the / night be/fore there was a / lot of uh ex–pec/ta–tion and ex/cite–ment on / my / part and um / af–ter / I went to to / sleep my / pa–rents would would / al–ways / o–pen / up um / se–ve–ral of the / gifts / se–ve–ral of the im/por–tant / all the sur/prise things and / so on / christ–mas / morn–ing you know / I would wake / up and go / run–ning in–to the / li–ving room and / it would be / filled with / all these / won–drous / things um / and / then there would be / some pre–sents that / weren't / o–pened / ty–pi–cal–ly the the / pre–sents that had / been / un–der the / tree be/fore / a–ny–way and / we would / o–pen / all of / those and / I re–mem–ber just / be–ing just a / v a / ve–ry a v a / real–ly / su–per / hap–py / kind of / time and / we would / leave / things / spread out all / o–ver the k the / li–ving room / floor and / things in a / mess / for / you know at / least the en at / least all / through that / day / some–times / se–ve–ral / days and / that was / kind of / spe–cial we / did–n't have to / clean / things / up and it was / good the / most the / thing I / think that / comes / clo–sest to a / fam–i–ly / ri–tu–al in / my / fam–i–ly / was uh / vi–sits to um my / grand–par–ents who / lived um / four hun–dred / miles a/way in um new / me–xi–co and / that's / real–ly the / only / trips we / e–ver took / a–ny–where I mean my / fam–i–ly did / not take va/ca–tions but / we would / go / vi–sit / gran–ma and / gran–pa / reed um / two / three / times a / year I / guess / so / there was that / eight ho–ur / au–to trip I / knew the / road / ve–ry / well and uh / they / lived in new / me–xi–co in / ve–ry uh sort of / se–mi / ar–e–a / a–rid / coun–try / san–dy / and so it was a / ve–ry / dif–fe–rent / kind of / world / there and / they / al–ways / trea–ted me / won–der–ful–ly my / grand–fa–ther / worked / out in the / oil fields / he was a / va–ri–ous / kind of / sales–man at / diff–er–ent / times and / he would / take me / with him when he went and / so / we would / tra–vel / I don't know / two hun–dred / miles a / day or / some–thing he would / tra–vel a/round / this and / that

Figure 3. The 'real' foot boundaries - majority votes of subjects H, A, and C.

2.2.2 Foot boundary uncertainty

As part of the foot boundary task, the subjects were directed to underline those portions of the text where they had difficulty marking boundaries and were uncertain.of themselves on the last two trials. The results were not particularly revealing, but I give them here for completeness. First the correlations[6] between the two sets of uncertainty markings and the sites at which the subject was demonstrably uncertain, that is, those sites marked on two, three, or four of the trials only. U5 and U6 label the uncertainty markings from trials five and six for each subject, and SS (Shaky Sites) is for the inconsistently marked sites. Subject D failed to mark uncertainty on trial 6.

Subject A				**Subject B**			
	Trial				Trial		
Trial		U5	U6	Trial		U5	U6
	U6	.204			U6	.215	
	SS	.237	.407		SS	.250	.234
Subject C							
	Trial			Subject D			
Trial		U5	U6		Trial		
	U6	.348		Trial		U5	
	SS	.437	.477		SS	.196	

Table 1. Correlation of uncertainty markings with sites of inconsistency

There is not too much to say about this - there is some correlation, but not much, except in the case of subject C, who is considerably more on target than the others.

The next table compares the uncertainty marking across subjects, and also looks at the uncertainty markings compared to the sites marked inconsistently across the summary trials of all five subjects. UA, UB, and UC labels the uncertainty marked on trial six by subjects A, B, and C; UD labels the uncertainity marked on trial five by subject D, and ShakySites labels those sites marked only two or three times in the summary trials of the five subjects.

[6] The measure here and subsequently in this section really is the correlation, not percentage agreement as in the previous section. Thus the null hypothesis is 0, with negative values showing inverse correlation.

	Subject			
Subject	UA	UB	UC	UD
UB	.022			
UC	-.049	.006		
UD	-.046	.068	-.010	
SS	.200	.099	.084	.168

Table 2. Correlation of final uncertainty markings and sites of inconsistency across summary trials

Clearly there is no correlation between the different subjects on the uncertainty marking, and little prediction of the sticky spots in the summary.

Finally, I checked the correlation of each subject's final uncertainty with the collection of sites where that subject's summary trial disagreed with the 'real' foot boundaries as given in Figure 3 in section 2.2.1 above, but there was not much here either.

Subject	A	B	C	D
Cor.	.116	.190	.215	.102

Table 3. Correlations of each subject's final uncertainity markings and sites of disagreement between that subject and the 'truth'.

The uncertainity marking seemed like a good idea when I was designing the experiment, but after the fact I have not come up with any good use to put the results to. Part of the problem was that using underlining to indicate uncertainity produced somewhat vague data. The lack of correlation between different subjects is due in part to differences in how particular each subject was in his/her underlining. Some thoughts on improving the utility of this part of the task will be given in section 2.3 below.

2.2.3 Tone group boundary assignment

The agreement matrices for the marking of tone group boundaries are given on the next page. Note that as in the previous task I did one less trial than the other subjects, and that the last trial (#5 for me, #6 for the others) is the summary trial.

On the page after that are graphs comparing the main diagonals and bottom rows of the agreement matrices.

Subject H 51 sites at issue, Null hypothesis = .055

	Trial			
Trial	1	2	3	4
2	.902			
3	.902	.961		
4	.882	.980	.980	
5	.882	.941	.902	.922

Mean = .922
Variance = .001

Subject A 57 sites at issue, Null hypothesis = .054

	Trial				
Trial	1	2	3	4	5
2	.912				
3	.825	.807			
4	.825	.807	.860		
5	.860	.842	.895	.895	
6	.825	.807	.860	.860	.965

Mean = .850
Variance = .002

Subject B 61 sites at issue, Null hypothesis = .058

	Trial				
Trial	1	2	3	4	5
2	.738				
3	.656	.885			
4	.754	.852	.902		
5	.705	.836	.885	.918	
6	.705	.803	.852	.951	.902

Mean = .814
Variance = .008

Subject C 46 sites at issue, Null hypothesis = .051

	Trial				
Trial	1	2	3	4	5
2	.935				
3	.957	.978			
4	.913	.978	.957		
5	.870	.935	.913	.913	
6	.913	.978	.957	1.000	.913

Mean = .941
Variance = .001

Subject D 50 sites at issue, Null hypothesis = .048

	Trial				
Trial	1	2	3	4	5
2	.880				
3	.800	.800			
4	.820	.860	.820		
5	.820	.900	.900	.880	
6	.760	.840	.880	.860	.940

Mean = .848
Variance = .002

Table 1. Percentage agreement between trials of tone group boundary placement, for each subject, relative to the sites at issue for that subject.

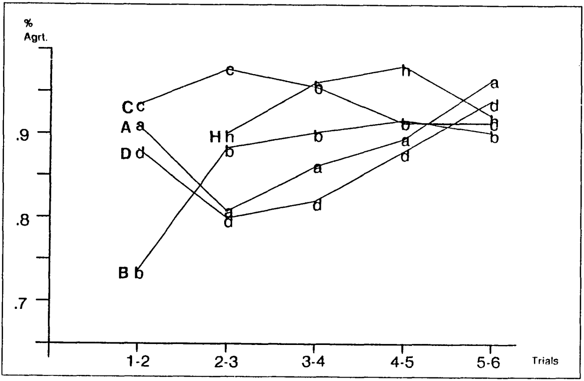


Figure 1a. Learning curves for Table 1, main diagonal (one trial vs. the next).

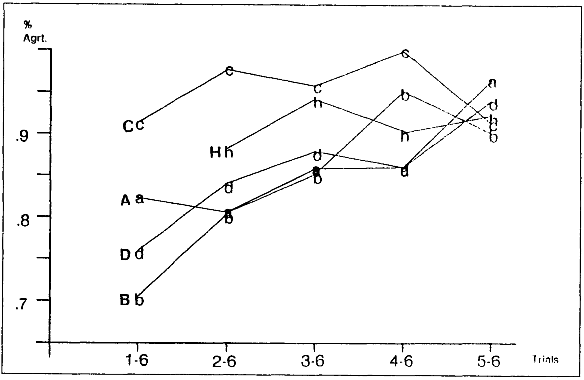


Figure 1b. Learning curves for Table 1, bottom row (each trial vs. summary).

The agreement matrices in Table 1 look quite good. This seems to have been an easier task than the previous one. Except for subject B, whose first trial is noticeably worse than the rest, the learning curves are not as steep as they were for the previous task, suggesting that the phenomenon was easily grasped from the instructions and training segment, and from prior experience. Subject C does extremely well here.

Restricting the statistics to the sites at issue has an even stronger effect here than it did for the previous task, as the number of sites at issue is much smaller. Once again, for comparison, the matrix below gives the agreement figures for subject A if all 451 sites are considered. Comparing this to the corresponding figures in Table 1, the inflation in the results is again apparent, ranging from 17 percentage points at the low end to 3 percentage points at the high end. The differential aspect of the inflation is even worse than for the previous task: The range of the data is compressed from 16 percentage points to 2 percentage points separating the highest from the lowest, and again the variance has dropped below .0005. Note that this compression results from the extremely high null hypothesis, which in turn results from all the spurious, no mark matching no mark, sites.

Subject A all 451 sites at issue, null hypothesis .816

	Trial				
Trial	1	2	3	4	5
2	.989				
3	.978	.976			
4	.978	.976	.982		
5	.982	.980	.987	.987	
6	.978	.976	.982	.982	.996

Mean = .981
Variance = .000

Table 2. Percentage agreement between trials of tone group boundary placement, for subject A, relative to all 451 sites.

The next table shows the number of tone group boundaries marked by each subject in each trial. Here again we see much greater stability than in the previous task, with the variances being significantly less. Again subject C shows up particularly well.

	Trial							
Subject	1	2	3	4	5	6	mean	var
H	44	47	47	48	48		46.800	2.7
A	40	39	50	50	48	50	46.167	27.4
B	43	45	52	52	55	53	49.667	21.5
C	44	45	46	44	42	44	44.167	1.8
D	37	41	41	42	42	43	41.000	4.4

Grand mean = 45.517, variance = 19.044

Mean of summary trials = 47.400, variance = 14.800

Table 3. Total number of tone group boundaries marked.

The next three tables show the consistency of tone group boundary marking. Table 4a gives the raw numbers, 4b and 4c give collapsed percentages. The nature of the collapsing and the labeling are as in the parallel tables in the previous section.

	Number of times tone group boundary marked						
Subject	0	1	2	3	4	5	6
H	400	3	1	2	2	43	
A	394	8	2	2	3	5	37
B	390	6	2	2	6	10	35
C	405	0	1	0	1	5	39
D	401	6	1	3	4	3	33

Table 4a. Number of sites receiving a given number of marks over all trials

	Consistency of tone group boundary marking			
Subject	Perfect	1 out	2 out	Worst
H	.843	.098	.059	
A	.649	.228	.088	.035
B	.574	.262	.131	.033
C	.848	.109	.043	0.000
D	.660	.180	.100	.060

Table 4b. Percentage of sites at issue receiving a given number of inconsistent marks over all trials

	Consistency of tone group boundary marking	
Subject	1 or less	2 or more
H	.941	.059
A	.877	.123
B	.836	.164
C	.957	.043
D	.840	.160

Table 4c. Collapsed percentage of sites at issue receiving a given number of inconsistent marks over all trials

These figures confirm the improved consistency of the subjects on this task as compared to the previous one, and subject C continues to shine.

Now we turn to the agreement across subjects. Tables 5a and 5b are the agreement matrices for the last or summary trial of each subject. They differ in terms of what the sites at issue were. Table 5a is based on the sites at issue over all 29 trials of the tone group boundary marking task. If any subject marked a site on any trial, then it is one of the sites at issue for this table. Table 5b takes a more conservative approach, considering only those sites marked in one of the five summary trials themselves.

Subject	69 sites at issue, Null hypothesis = .053			
Trial				
Trial	H	A	B	C
A	.855			
B	.811	.811		
C	.855	.855	.840	
D	.812	.870	.797	.841

Mean = .835
Variance = .001

Table 5a. Percentage agreement of summary tone group boundary placements across subjects, sites at issue over all trials

Subject	59 sites at issue, Null hypothesis = .055				
Trial					
Trial	H	A	B	C	
A	.831				
B	.780	.780			
C	.831	.831	.814		
D	.780	.847	.763	.814	
					Mean = .807
					Variance = .001

Table 5b. Percentage agreement of summary tone group boundary placements across subjects,sites at issue over summary trials

These figures show that not only did subjects agree more with themselves, they also agreed more generally with each other. The agreements for subjects H, A, and C are quite similar to the parallel figures for the previous task, but this time subjects B and D are much more in line with their peers.

This agreement is also manifest in the next table, which shows the absolute and percentage consistency figures for the summary trials.

	Number of times tone group boundary marked					
	0	1	2	3	4	5
	392	5	8	3	7	36
%		8.5	13.6	5.1	11.9	61.0

Table 6. Number of sites receiving a given number of marks over the five summary trials.

Over 60% of the sites at issue were marked unanimously, an improvement of more than 20 percentage points from the previous task. There does not appear to be a subset of the subjects which would significantly improve these figures as is the case for the previous task.

The union of all the summary trials is displayed below. Plain vertical bars (|) indicate a unanimous marking - those sites with a less than unanimous marking are marked with a vertical bar with the subjects marking at that site indicated by letter. For example, $^{a}|^{c}$ stands for a site marked by subjects A and C only.

I know that ${}^{a}_{d}|^{b}$ um on the night be–fore ${}^{b}|^{c}$ there was a lot of uh ex–pec–ta–tion and ex–cite–ment on my part | and um af–ter I went to to sleep | my pa–rents would would al–ways o–pen up ${}^{a}|^{b}$ um se–ve–ral of the gifts | se–ve–ral of the im–por–tant ${}^{a}_{d}|^{b}$ all the sur–prise things | and so on christ–mas morn–ing ${}^{b}|^{c}$ you know ${}^{b}|$ I would wake up | and ${}^{a}|$ go run–ning in–to the li–ving room | and it would be filled with all these won–drous things | um ${}^{b}|$ and ${}^{h}|^{b}$ then ^{hb} | there would be some pre–sents that weren't o–pened | ty–pi–cal–ly the the pre–sents that had been un–der the tree be–fore a–ny–way | and we would o–pen all of those | and I re–mem–ber just be–ing just a v a ve–ry ${}^{a}|^{d}$ a v a real–ly su–per hap–py kind of time | and we would leave things spread out all o–ver the k the li–ving room floor ${}^{h}_{c}|^{a}_{d}$ and ${}^{b}|$ things in a mess | for you know at least the en at least all through that day | some–times se–ve–ral days | and that was kind of spe–cial | we did–n't have to clean things up | and it was good | the most ${}^{a}_{c}|^{b}_{d}$ the thing I think that comes clo–sest to a fam–i–ly ri–tu–al | in my fam–i–ly | was uh vi–sits to um my grand–par–ents | who lived um four hun–dred miles a–way | in um new me–xi–co | and that's real–ly the only trips we e–ver took a–ny–where | I mean my fam–i–ly did not take va–ca–tions | but we would go vi–sit gran–ma and gran–pa reed | um two three times a year I guess | so ${}^{h}|$ there was that eight ho–ur au–to trip | I knew the road ve–ry well | and uh they lived in new me–xi–co ${}^{h}_{d}|^{b}$ in ve–ry uh sort of se–mi ar–e–a a–rid coun–try | san–dy | and so it was a ve–ry dif–fe–rent kind of world there | and they al–ways trea–ted me won–der–ful–ly | my grand–fa–ther worked out in the oil fields | he was a va–ri–ous kind of sales–man ^{ha} ${}_{b}|_{c}$ at diff–er–ent times | and he would take me with him ${}^{h}_{b}|^{a}_{d}$ when he went | and so ${}^{h}|^{a}$ we would tra–vel ${}^{h}_{b}|^{a}_{c}$ I don't know ${}^{h}_{b}|^{a}_{c}$ two hun–dred miles a day or some–thing ${}^{h}_{b}|^{a}_{c}$ he would tra–vel a–round ${}^{h}|^{a}$ this and that |

Figure 2. The union of tone group boundary marking of all subjects.

Various classes of marking are apparent in the foregoing: Unanimous marks, marks agreed on by three or four subjects, marks agreed on by two subjects which are associated with anomalies in the smooth flow of the text, marks made by one or two subjects associated with isolated words, marks made by only one subject which seem to be in error, and marks made by two subjects which seem to represent genuine disagreement about what is going on. The resolution of all these into a set of 'real' tone groups is attempted on the following page. The first two categories, being boundaries agreed on by three or more subjects, are all carried over, marked with a simple vertical bar (|). There were two sites in the next category, being marked by two subjects only, and located at rough spots in the text, at the beginning of lines 3 and 8. These have not been carried over. The next category, being sites marked by one or two subjects around isolated, drawn out, space filling words, are carried over, but marked with a double bar (||). In both these situations, the lack of unanimity presumably stems from the failure of the directions to specify how to deal with these sort of anomalous situations. In the next category are three sites, on lines 4, 5, and 9, which are marked only by one subject, one site away from a consensus site, which seem clearly in error and are not carried over. The remaining category covers four sites where there is no immediately obvious reason for the lack of unanimity, nor any principled reason for supposing they are simply in error. These are carried over, but marked with a section mark (§).

I know that | um on the night be–fore § there was a lot of uh ex–pec–ta–tion and ex–cite–ment on my part | and um af–ter I went to to sleep | my pa–rents would would al–ways o–pen up um se–ve–ral of the gifts | se–ve–ral of the im–por–tant | all the sur–prise things | and so on christ–mas morn–ing § you know § I would wake up | and go run–ning in–to the li–ving room | and it would be filled with all these won–drous things | um and || then || there would be some pre–sents that weren't o–pened | ty–pi–cal–ly the the pre–sents that had been un–der the tree be–fore a–ny–way | and we would o–pen all of those | and I re–mem–ber just be–ing just a v a ve–ry a v a real–ly su–per hap–py kind of time | and we would leave things spread out all o–ver the k the li–ving room floor | and things in a mess | for you know at least the en at least all through that day | some–times se–ve–ral days | and that was kind of spe–cial | we did–n't have to clean things up | and it was good | the most | the thing I think that comes clo–sest to a fam–i–ly ri–tu–al | in my fam–i–ly | was uh vi–sits to um my grand–par–ents | who lived um four hun–dred miles a–way | in um new me–xi–co | and that's real–ly the only trips we e–ver took a–ny–where | I mean my fam–i–ly did not take va–ca–tions | but we would go vi–sit gran–ma and gran–pa reed | um two three times a year I guess | so || there was that eight ho–ur au–to trip | I knew the road ve–ry well | and uh they lived in new me–xi–co | in ve–ry uh sort of se–mi ar–e–a a–rid coun–try | san–dy | and so it was a ve–ry dif–fe–rent kind of world there | and they al–ways trea–ted me won–der–ful–ly | my grand–fa–ther worked out in the oil fields | he was a va–ri–ous kind of sales–man | at diff–er–ent times | and he would take me with him | when he went | and so || we would tra–vel | I don't know | two hun–dred miles a day or some–thing | he would tra–vel a–round § this and that |

Figure 3. The 'real' tone group boundaries - majority votes of all subjects, with a few adjustments. § indicates a plausible non-majority site, || a 'defective' boundary - see discussion above.

2.2.4 Kinetic tone location and identification

The agreement matrices for the kinetic tone marking task are given below. As in the previous task I did one less trial than the other subjects, and that the last trial (#5 for me, #6 for the others) is the summary trial.

The form of the tables below is somewhat different from that of the previous sections. Because this task was not just an either/or question like the previous ones, but rather a question of categorizing, there are really two separable tasks being done: Deciding where in the data segment there are kinetic tones, and having done so deciding whether they are rising, falling, etc. So the entries in the tables below are paired. The first half gives the agreement for the first, or kinetic tone location, task, and the second gives the agreement for the second, or kinetic tone identification task.

Subject H		74 sites at issue, Null hypothesis = .056				
	Trial					
Trial		1	2	3	4	
	2	.784/.770				
	3	.676/.649	.784/.730			
	4	.757/.662	.811/.730	.811/.689		
	5	.473 /.338	.554/.419	.527/.392	.446/.351	
					Mean = .651/.557	
					Variance = .019/.028	
Subject A		68 sites at issue, Null hypothesis = .049				
	Trial					
Trial		1	2	3	4	5
	2	.809/.794				
	3	.544/.471	.588/.544			
	4	.691/.588	.706/.588	.676/.603		
	5	.676/.618	.779/.706	.721/.647	.779/.676	
	6	.735/.676	.838/.750	.662/.559	.691/.662	.853/.779
					Mean = .712/.637	
					Variance = .007/.007	
Subject B		96 sites at issue, Null hypothesis = .066				
	Trial					
Trial	1	2	3	4	5 2	.615/.458
	3	.469/.344	.687/.646			
	4	.448/.333	.687/.615	.708/.667		
	5	.500/.385	.698/.625	.740/.698	.677/.646	
	6	.521/.406	.656/.604	.740/.708	.740/.719	.896/.854
					Mean = .639/.566	
					Variance = .014/.024	
Subject C		74 sites at issue, Null hypothesis = .052				
	Trial					
Trial	1	2	3	4	5	
	2	.649/.514				
	3	.595/.500	.622/.514			
	4	.595/.486	.676/.568	.703/.608		
	5	.662/.527	.635/.527	.797/.649	.716/.622	
	6	.676/.581	.730/.608	.595/.527	.676/.595	.662/.554
					Mean = .665/.558	
					Variance = .003/.002	
Subject D		86 sites at issue, Null hypothesis = .068				
	Trial					
Trial	1	2	3	4	5	
	2	.651/.558				
	3	.640/.570	.640/.535			
	4	.651/.605	.674/.581	.826/.663		
	5	.767/.674	.674/.535	.733/.651	.674/.581	
	6	.593/.512	.640/.512	.791/.698	.709/.570	.756/.674
					Mean = .692/.593	
					Variance = .004/.004	

Table 1. Percentage agreement between trials of kinetic tone location and identification, for each subject, site per syllable version, relative to the sites at issue for that subject.

These numbers do not look particularly impressive. Compared to the previous two tasks, the overall means are not high, the variances are high, and there is little or no sign of improving across trials. All the subjects do seem to be better at kinetic tone location than at identification, by around 10 percentage points. Subject A is noticeably better than the others, but still not great.

Looking at the individual transcriptions reveals that at least some of the problem stemmed from inconsistency as to which of several adjacent syllables was marked from one trial to the next. The next set of agreement matrices attempts to see what would happen if that indeterminacy were removed, by collapsing the syllables of each word together into a single site, so that e.g. a fall on the first syllable of a word would agree with a fall marked on the second syllable of that word on a subsequent trial. In a few cases a subject would have marked more than one kinetic tone within a single word, but the vast majority of these cases were a rise and a fall, which were combined to a rise-fall, and a fall and a rise, which were combined to a fall-rise.

Subject H	59 sites at issue, Null hypothesis = .056				
	Trial				
Trial 1	2	3	4		
2	.797/.780				
3	.763/.729	.898/.831			
4	.763/.644	.898/.797	.864/.695		
5	.746/.492	.814/.559	.780/.542	.780/.593	
				Mean = .804/.651	
				Variance = .003/.013	

Subject A	56 sites at issue, Null hypothesis = .048				
	Trial				
Trial 1	2	3	4	5	
2	.875/.857				
3	.679/.571	.732/.625			
4	.750/.607	.768/.607	.750/.625		
5	.839/.732	.893/.768	.768/.661	.839/.714	
6	.875/.768	.964/.821	.768/.643	.804/.768	.929/.839
				Mean = .814/.702	
				Variance = .006/.008	

Subject B	78 sites at issue, Null hypothesis = .066				
	Trial				
Trial 1	2	3	4	5	
2	.718/.513				
3	.692/.449	.795/.718			
4	.564/.385	.744/.654	.769/.718		
5	.705/.500	.782/.667	.782/.731	.731/.692	
6	.679/.449	.756/.667	.756/.705	.782/.731	.949/.897
				Mean = .738/.616	
				Variance = .006/.019	

Subject C	54 sites at issue, Null hypothesis = .051				
	Trial				
Trial 1	2	3	4	5	
2	.778/.500				
3	.741/.593	.852/.611			
4	.741/.574	.741/.519	.852/.704		
5	.759/.500	.796/.519	.907/.648	.796/.611	
6	.778/.593	.815/.611	.852/.722	.852/.685	.796/.574
				Mean = .801/.597	
				Variance = .003/.005	

Subject D	67 sites at issue, Null hypothesis = .059				
	Trial				
Trial 1	2	3	4	5	
2	.716/.552				
3	.716/.567	.642/.478			
4	.716/.597	.731/.552	.881/.642		
5	.746/.612	.731/.507	.821/.657	.761/.582	
6	.687/.507	.701/.493	.881/.687	.791/.612	.851/.672
				Mean = .756/.580	
				Variance = .004/.004	

Table 2. Percentage agreement between trials of kinetic tone location and identification, for each subject, site per word version, relative to the sites at issue for that subject.

This change improves things considerably. The number of sites at issue drops, the means go up, the variances go down. My summary trial is still clearly a disaster. H and C were improved the most, while D was not helped much. A no longer stands out so much. But there is still not much sign of learning, although the location agreement percentages in the high 80% range are almost as good as those in the tone group boundary task, and in some cases better than those in the foot boundary task.

The kinetic tone identification numbers still look quite bad, however, not being improved by the change as much as the location numbers.

I tried a further collapse, reducing all the syllables within a foot to a single site, but this did not yield significant further improvements in the agreement figures, with e.g. the kinetic tone location means only increasing by between 1 and 4 percentage points.

The next table show the number of sites at which kinetic tones were marked by each subject on each trial.

	Trial							
Subject	1	2	3	4	5	6	mean	var
H	43	49	49	51	48		48.000	9.0
A	41	38	50	42	39	41	41.833	18.167
B	52	49	57	61	56	58	55.500	18.700
C	43	45	45	43	46	45	44.500	1.500
D	42	58	59	62	56	65	57./50./47.	64.000

Grand mean = 49.414, variance = 57.680

Mean of summary trials = 51.400, variance = 97.300

Table 4. Total number of sites marked with some kinetic tone

These figures are for the site per syllable version of the data, but are not significantly different from those for the site per word or site per foot versions, except for subject D. The three figures in the mean column for subject D are for the site per syllable, per word, and per foot versions respectively, and show that a significant difference did occur for this subject. This resulted from a greater tendency to mark e.g. a rise and a fall on two syllables of a disyllabic word rather than simply a rise-fall on one of those syllables, which lead to a difference in by syllable versus by word site counts.

The next tables are similar to the consistency tables given for the previous two tasks, although the interpretation is somewhat different owing to the dual nature of this task. Thus e.g. the value 14 in the third column of the first row for subject A means that there were 14 sites at which a particular kinetic tone was marked exactly twice out of the six trials by subject A. This figure is the total for all four possible kinetic tone types. None the less the overall interpretation is the same as for the previous tables of this type: The 6 column is good and the 3 column is bad. The 0 column is not particularly informative, as e.g. a site which was marked \ \ \ V V V would score as a site where both V and \ were marked three times, *and* where both / and ∧ were marked zero times. The figures are given for all three version of the data.

	Number of times kinetic tone identified						
Subject	0	1	2	3	4	5	6
H by syll	205	29	15	20	14	13	
by word	158	18	12	15	12	21	
by foot	152	17	11	14	11	23	
A	187	31	14	9	8	5	18
	151	23	10	6	6	8	20
	148	22	11	5	7	7	20
B	262	45	25	11	11	19	11
	207	32	18	8	10	22	15
	204	32	16	8	10	24	14
C	193	39	22	14	9	8	11
	126	34	12	9	12	10	13
	117	33	9	8	13	10	14
D	229	36	18	23	7	13	18
	165	34	15	20	8	13	13
	148	33	15	20	8	14	10

Table 4a. Number of sites at which a kinetic tone was identified a given number of times.

The next table is a percentage version of the last three columns of Table 4a, covering the sites where a given kinetic tone received a majority of the marks. The last column is the total of the previous three, and gives a sort of overall consistency measure.

	Votes				
Subject	3	4	5	6	total
H by syll	.270	.189	.176		.635
by word	.254	.203	.356		.814
by foot	.246	.193	.404		.842
A		.118	.074	.265	.457
		.107	.143	.357	.607
		.127	.127	.364	.618
B		.115	.198	.115	.427
		.128	.282	.192	.603
		.130	.312	.182	.623
C		.122	.108	.149	.378
		.222	.185	.241	.648
		.255	.196	.275	.725
D		.081	.151	.209	.442
		.119	.194	.194	.507
		.129	.226	.161	.516

Table 4b. Percentage of sites at issue receiving a given number of kinetic tone identifications

Both the preceding tables reinforce the conclusions that the kinetic tone identification task was not easy, and that the reduction to a site per word basis improves the consistency, but that the further reduction to a site per foot basis does not improve things comparably. It is interesting to note that although subject H has the best total values, subject A also looks good if we consider the 6 column, which represents percentage of sites where a unanimous judgment was made on all six trials. Compared to the parallel tables for the previous two tasks, once again this task comes up the clear loser.

Next we consider the agreement across subjects of the summary trials, where once again the paired entries are for the location and identification subtasks, and we consider both the site per syllable and site per word versions.

95 sites at issue, Null hypothesis = .061

	Subject			
Subject	H	A	B	C
A	.695/.632			
B	.747/.674	.568/.495		
C	.526/.368	.558/.421	.484/.295	
D	.526/.358	.621/.421	.547/.400	.579/.379
				Mean = .586/.443
				Variance = .006/.013

Table 5a. Percentage agreement between summary trials of kinetic tone location and identification, across subjects, site per syllable version, sites at issue over all summary trials

77 sites at issue, Null hypothesis = .058

	Subject			
Subject	H	A	B	C
A	.792/.675			
B	.740/.649	.636/.481		
C	.571/.364	.675/.494	.571/.312	
D	.662/.416	.688/.390	.714/.506	.675/.351
				Mean = .674/.466
				Variance = .004/.013

Table 5b. Percentage agreement between summary trials of kinetic tone location and identification, across subjects, site per word version, sites at issue over all summary trials

Again, mixed reviews. The kinetic tone location numbers in the site per word version are just about up to the comparable results for foot boundaries, but the kinetic tone identification numbers are no good at all. Subjects H, A, and B seem a bit more in tune than the other two, but not by much.

The consistency tables across the summary trials show a similar pattern:

	Number of times kinetic tone identified					
version	0	1	2	3	4	5
by syll	239	72	38	19	8	4
by word	185	56	34	15	14	4
by foot	167	50	33	15	13	6

Table 6a. Number of sites at which a kinetic tone was identified a given number of times, summary trials for all subjects.

	Votes			
version	3	4	5	total
by syll	.200	.084	.042	.326
by word	.195	.182	.052	.429
by foot	.211	.183	.085	.479

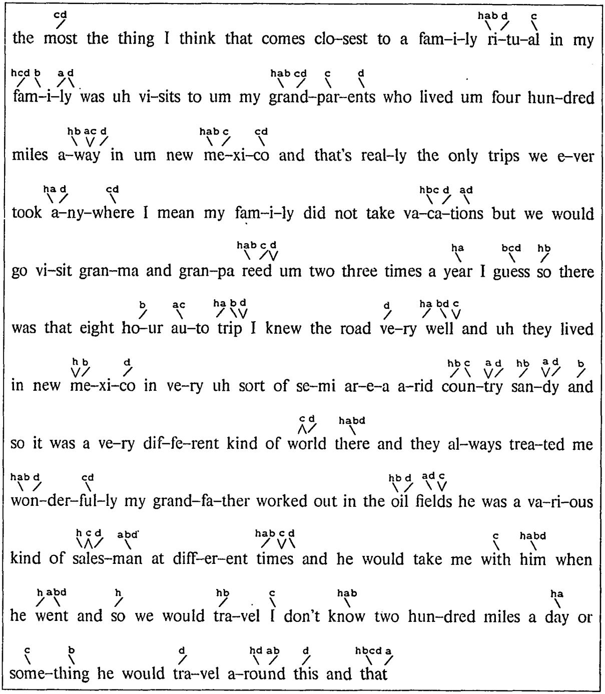
Table 6b. Percentage of sites at issue receiving a given number of kinetic tone identifications, summary trials for all subjects

The very low values in the unanimous columns are noticeable here - the subjects just don't seem to have a common model of this task.

Finally, the next figure gives a graphic demonstrate of the confused state of things, being a summary of the summary trials, similar in style to the ones given in the previous section. The symbols above the syllables indicate which marks were made at that site, and the letters above the symbols show which subjects made which marks.

```
         b                     bd c                              bd
         /                     /Λ                                /
I know that um on the night be-fore there was a lot of uh ex-pec-ta-tion and

   bd              ac bd                           hbcd
   /               V /                               /
ex-cite-ment on my part and um af-ter I went to to sleep my pa-rents would

                    hbd                                                  a bd
                    /                       \                            \ /
would al-ways o-pen up um se-ve-ral of the gifts se-ve-ral of the im-por-tant

          c       habd               d         d   bc d d   b
          \        \                 /         \   \ /  \   /
all the sur-prise things and so on christ-mas morn-ing you know I would wake

hbd c                        h       bd c
/ \                          \       / \
up and go run-ning in-to the li-ving room and it would be filled with all

       d        habd c b   hb  hb
       /         \ Λ   /   /   /
these won-drous things um and then there would be some pre-sents that weren't

b  hcd a
/  / V
o-pened ty-pi-cal-ly the the pre-sents that had been un-der the tree be-fore

hab  ad c                          hbd ac
\    / V                            / V
a-ny-way and we would o-pen all of those and I re-mem-ber just be-ing just a

                                             hab cd                    b
                                              \ Λ                      /
v a ve-ry a v a real-ly su-per hap-py kind of time and we would leave things

                                            h bd c                habd c
                                            \/Λ                    \ Λ
spread out all o-ver the k the li-ving room floor and things in a mess for you

                              c                habd c
                              \                 \ Λ                      \
know at least the en at least all through that day some-times se-ve-ral days and

                hb d acd                     d
                 \/   \                      /            \              \
that was kind of spe-cial we did-n't have to clean things up and it was good
```

Figure 3a. First part of summary of the kinetic tone markings from the summary trials of all subjects



the most the thing I think that comes clo–sest to a fam–i–ly ri–tu–al in my
fam–i–ly was uh vi–sits to um my grand–par–ents who lived um four hun–dred
miles a–way in um new me–xi–co and that's real–ly the only trips we e–ver
took a–ny–where I mean my fam–i–ly did not take va–ca–tions but we would
go vi–sit gran–ma and gran–pa reed um two three times a year I guess so there
was that eight ho–ur au–to trip I knew the road ve–ry well and uh they lived
in new me–xi–co in ve–ry uh sort of se–mi ar–e–a a–rid coun–try san–dy and
so it was a ve–ry dif–fe–rent kind of world there and they al–ways trea–ted me
won–der–ful–ly my grand–fa–ther worked out in the oil fields he was a va–ri–ous
kind of sales–man at diff–er–ent times and he would take me with him when
he went and so we would tra–vel I don't know two hun–dred miles a day or
some–thing he would tra–vel a–round this and that

Figure 3b. Second part of summary of the kinetic tone markings from the summary trials of all subjects

It seems pointless given the above confusion to try to construct a consensus marking of the data segment for kinetic tone, although a reasonable job could probably be done for location, if not for identification. It is clear that more extensive and structured training in the identification of tones is required to achieve agreement across subjects. John Trim has produced training materials of this sort, consisting of cassettes with accompanying transcriptions and exercises: It would be interesting to test the efficacy of these materials by repeating the kinetic tone marking task with subjects who have used them.

2.2.5 Tonal excursion location and identification

In presenting the results for the tonal excursion task, I have reversed the order of presentation of the first two tables from the previous sections, to show that there were serious problems with this task, at least for this text. Table 1 below gives the number of tonal excursions marked by each subject in each trial, whether upwards or downwards.

	Trial							
Subject	1	2	3	4	5	6	mean	var
H	19	13	10	12	15		13.800	11.7
A	13	14	14	9	6	8	10.667	11.9
B	17	9	10	9	11	12	11.333	9.1
C	11	2	8	3	9	11	7.333	15.5
D	10	14	10	7	8	4	8.833	11.4
	Grand mean = 10.276, variance = 14.993							

Table 1. Total number of sites marked with some tonal excursion.

The picture here is of few excursions marked, and high variances reflecting the considerable swings up and down in the number of sites marked. My own sense from doing the task was that, at least for the particular text being transcribed, that it sort of fell through the cracks. There were not enough extreme excursions to keep one alert for the phenomenon. More on this at the conclusion of this section.

The next table gives the agreement matrices for the task. There was little or no confusion of upward and downward excursions, in fact almost no downward excursions were marked, so the distinction between the two subtasks of location and identification which was made in the previous section is not made here. Also the locality problem of the previous task did not appear here, so no compression to a site per word version of the data is given either.

Subject H 26 sites at issue, Null hypothesis = .016

	Trial			
Trial	1	2	3	4
2	.538			
3	.654	.731		
4	.731	.731	.846	
5	.538	.538	.654	.731

Mean = .659
Variance = .009

Subject A 23 sites at issue, Null hypothesis = .012

	Trial				
Trial	1	2	3	4	5
2	.870				
3	.609	.478			
4	.565	.609	.348		
5	.435	.391	.478	.435	
6	.609	.565	.652	.435	.826

Mean = .546
Variance = .016

Subject B 26 sites at issue, Null hypothesis = .013

	Trial				
Trial	1	2	3	4	5
2	.346				
3	.346	.692			
4	.462	.692	.615		
5	.462	.615	.615	.692	
6	.500	.654	.577	.808	.808

Mean = .585
Variance = .019

Subject C 24 sites at issue, Null hypothesis = .008

	Trial				
Trial	1	2	3	4	5
2	.625				
3	.625	.750			
4	.583	.875	.708		
5	.417	.708	.458	.583	
6	.500	.625	.625	.583	.417

Mean = .603
Variance = .016

Subject D 23 sites at issue, Null hypothesis = .010

	Trial				
Trial	1	2	3	4	5
2	.565				
3	.522	.435			
4	.435	.435	.783		
5	.478	.304	.609	.652	
6	.565	.478	.696	.739	.739

Mean = .548
Variance = .018

Table 2. Percentage agreement between trials of tonal excursion assignment, for each subject, relative to the sites at issue for that subject.

Low means and high variances, and a lack of any substantial improvement over time characterize these numbers, with the exception of subject B, who seems to have a more coherent approach, with less oscillation in the number of sites marked, and a noticeable learning trend, and to a lesser extent subject H, whose overall mean is pretty good, but who shows no particular improvement over trials.

The next tables give the consistency figures.

	Number of times tonal excursion assigned						
Subject	0	1	2	3	4	5	6
H	78	11	4	1	3	7	
A	69	6	5	5	3	3	1
B	74	17	1	5	2	4	1
C	70	17	5	1	1	2	0
D	67	12	5	4	2	1	1

Table 3a. Number of sites at which a tonal excursion was identified a given number of times

	Votes				
Subject	3	4	5	6	Total
H	.038	.115	.269		.423
A		.130	.130	.043	.304
B		.077	.154	.038	.269
C		.042	.083	0.0	.167
D		.087	.043	.043	.174

Table 3b. Percentage of sites at issue receiving a given number of tonal excursion identifications

These are far and away the worst numbers for these statistics we have yet seen. Subject H is the only one with more than one site marked completely consistently, and subject C did not mark even *one* site completely consistently. The totals, being the percentage of sites where a particular marking received a majority of the votes, are considerably less than for the previous tasks.

The tables comparing the summary trials across subjects are also pretty bad. Again we start with the table for the number of sites marked, revealing a substantial range:

Subject						
H	A	B	C	D	mean	var
15	8	12	11	4	10.000	17.5

Table 4. Total number of sites marked with some tonal excursion across summary trials

With this much disagreement it is not surprising that the agreement matrix also looks pretty grim:

26 sites at issue, Null hypothesis = .011

	Subject			
Subject	H	A	B	C
A	.654			
B	.692	.731		
C	.308	.423	.346	
D	.423	.615	.423	.500

Mean = .500
Variance = .021

Table 5. Percentage agreement between summary trials of tonal excursion assignment, across subjects, sites at issue over all summary trials

And the consistency of voting across the summary trials is also dismal:

	Votes				
	0	1	2	3	4
	77	14	5	6	2
%		53.8	19.2	23.1	7.7

Table 6. Number of sites at which a tonal excursion was identified a given number of times, summary trials for all subjects.

Slightly less than 31 percent of the sites were marked by even a majority of the five subjects, and almost 54% were marked by only one subject. There were *no* sites marked unanimously.

The final, graphic evidence for the lack of coherence on this task is given by the following summary presentation of the tonal excursion markings from the summary trials, in the same style as for the previous tasks.

```
                                                               c
                                                               ⊤
I know that um on the night before there was a lot of uh expectation and
   c
   ⊤
excitement on my part and um after I went to to sleep my parents would would
        b                                               d       hbc       had b
        ⊤                                               ⊥        ⊤         ⊤  ⊥
always open up um several of the gifts several of the important all the surprise
things and so on christmas morning you know I would wake up and go running
         h                                                          hab
         ⊤                                                           ⊤
into the living room and it would be filled with all these wondrous things um
                                                          c
                                                          ⊤
and then there would be some presents that weren't opened typically the the
                                        h    hd
                                        ⊤     ⊤
presents that had been under the tree before anyway and we would open all of
         hab
          ⊤
those and I remember just being just a v a very a v a really super happy kind
  hab
    ⊤
of time and we would leave things spread out all over the k the living room
                                           habc            habc
                                             ⊤               ⊤
floor and things in a mess for you know at least the en at least all through
that day sometimes several days and that was kind of special we didn't have to
                                    hab        b
                                     ⊤         ⊤
clean things up and it was good the most the thing I think that comes closest
            hc
            ⊤
to a family ritual in my family was uh visits to um my grandparents who lived
                                                 hb     a
                                                 ⊤      ⊤
um four hundred miles away in um new mexico and that's really the only trips
             cd
             ⊤
we ever took anywhere I mean my family did not take vacations but we would
                           hb                          c
                           ⊤                           ⊤
go visit granma and granpa reed um two three times a year I guess so there
was that eight hour auto trip I knew the road very well and uh they lived in
              h
              ⊤
new mexico in very uh sort of semi area arid country sandy and so it was a
                                                               c
                                                               ⊤
very different kind of world there and they always treated me wonderfully my
grandfather worked out in the oil fields he was a various kind of salesman at
different times and he would take me with him when he went and so we would
                                         c
                                         ⊤
travel I don't know two hundred miles a day or something he would travel
around this and that
```

Figure 1. Summary of the tonal excursion markings from the summary trials of all subjects.

It is clear from this figure, together with all the preceding tables, that this task was not successful, although there are a few sites with substantial agreement, where there is clearly something going on. Things are actually worse than the agreement matrix across subjects shows, because the number of sites at issue is high compared to the number of sites marked on any given trial. This inflates the agreement figures somewhat, because the validity of the approach I have taken to computing these numbers depends on there being a reasonable agreement between the set of sites at issue for each of the subjects, as has bee.ı pretty much the case for the other tasks. The issue of why the task differed so much in the level of their results will be considered further in section 2.3 below.

2.2.6 Intonational uncertainty

I was not able to make any significant use of the uncertainity markings on the 'b' worksheets. Different subjects seemed to use the underlining in different ways, and the results for the kinetic tone and tonal excursion tasks were so sloppy in any case that it was impossible to make any sensible comparison between loci of disagreement and uncertainty markings.

2.3 General discussion

To introduce this concluding discussion, one last table. The entries in this table for subjects against tasks are as close to a distillation of the performance of each subject on each task into a single number as I can get. They are a measure of how close the subject came to being perfectly consistent on each task, weighted in favor of learning and stability of judgment. At each site, the subject is scored on the basis of the number of trials it is possible to count back from the *last* trial without reaching a conflicting mark. Thus a parking pattern of e.g. 0 0 1 1 1 1 would score 4, where as a site with the same number of marks, but in a different pattern would score less, e.g. 1 0 1 0 1 1 would score only 2, as would 1 1 1 1 0 0. Another way to look at it is that a subject's score at a site is determined by the number of trials there were after he changed his mind for the last time. The single figure given below is obtained by summing over all the sites at issue for the subject, and then dividing by the perfect score, which would be a 6 at every site, or 6 times the number of sites at issue. Row and column, that is by task and by subject, means are given for this test as well.

The column labelled *summary* gives the result of a similarly conceived test across the summary trials for each task. Since there is no order to these trials as there is for the within subject trials, a different method of scoring each site is used: The score is the margin by which the mark receiving a plurality of the votes beat the combined totals of the opposition. For example, for a site marked 0 0 1 0 1 the score is 1, since three zeros are one more than two ones. For the two multivalued tasks, the score at a site can be negative, e.g. at a site marked \ \ ∕ 0 V the score would be –1, as two falls has the plurality, but there are three votes against that marking. As for the other previous test, the single number given is the ratio of the score obtained to the best possible score, in this case five times the number of sites at issue.

Subject							
Task	H	A	B	C	D	mean	summary
foot bnd	.871	.796	.789	.791	.619	.773	.681
tg bnd	.945	.848	.842	.938	.850	.885	.769
kin tone	.400	.559	.606	.505	.529	.520	.309
tnl excr	.608	.399	.538	.396	.493	.487	.431
mean	.706	.650	.694	.657	.623		

Table 1. Grand summary test statistics across subjects and trials

This table presents in a nutshell many of the general conclusions stated throughout the preceding section. On the foot boundary task, subject H is noticeably better and subject D noticeably worse than the others. The tone group boundary task is far and away the most solid, both for the subjects individually and overall. As compared to the two temporal tasks, the two intonational tasks do not look very good. For the kinetic tone task, the problem seems to be mostly one of accurately and consistently distinguishing falls from rises, and so on. There is some hope that performance here can be improved by more extensive training on just this problem. But the problem with the tonal excursion task seems to go deeper. It may be that the data segment itself is partly to blame, in that being a relatively calm monologue, not too much use is being made by the speaker of strongly marked tonal prominences.

But it is also clear that there are problems with my approach to the phenomenon. I have attempted to put an either/or straitjacket on what is most probably a gradient of prominence. My subjective experience in doing the task was that although it was clear to me that there was a certain small amount of tonal prominence associated with most foot boundaries, and there were a few overwhelmingly obvious peaks (although not so overwhelming that all the other subjects agreed!), it was not at all clear how I was supposed to divide up the intermediate cases. I had hoped to appeal to a categorization based less on the immediate tonal perception, but on some subjective experience of prominence, but it is clear that even supposing such an experience exists and can be consciously accessed, my directions and the structure of the task were such that it did not emerge in the results. Perhaps a task couched in more subjective terms, such as "Underline the words in this text which stand out to you", *not* in the context of other, acoustically based tasks, would yield better results, but I am not as confident of this as I would like to be.

On the other hand it is precisely because the results for the temporal tasks look so good that we can treat the negative results seriously. The fact that these percepts are stable and shared is encouraging, and at least these aspects of the framework proposed in Chapter 1 seem to have been strongly confirmed by the experiment. The next two chapters take the temporal categorizations established here and build on them, both in further investigation of phenomenal aspects of the data segment, in Chapter 3, and as a basis for theorizing at the functional level, in Chapter 4.

Chapter 3. Objective properties of feet: The duration of syllables

One of the results of the transcription experiment described in the previous chapter was a division of the data segment into feet, which represented the consensus of three of the subjects and which we judged to be close to the 'truth'. This chapter reports on the results of examining the distribution of the duration of the syllables of the data segment with respect to their location within those feet. These results are not consistent with a strict interpretation of the isochronicity hypothesis, but are consistent with an alternative hypothesis, which will be set forth.

3.1 Temporal aspects of the data segment

Using the system described in Appendix C, which enabled me to examine the time-amplitude acoustic waveform of the data segment, and to visually establish and aurally confirm the division of the data into words and syllables, I obtained a complete analysis of the duration of the syllables and filled and unfilled pauses of the data segment. The accuracy of this segmentation is quite good: I would say that most of the boundaries are accurate to plus or minus one glottal wave, which is about 10 milliseconds, as the fundamental frequency of the speaker, an adult male, varies around 100 hertz. I had some difficulty distinguishing unvoiced fricatives from silence, especially at the end of words, as a result of characteristics of the system employed (see Appendix C), but this did not have a significant effect over all.

The entire data segment is 130.20 seconds long, of which 51.03 seconds, or 39%, is taken up by filled or unfilled pauses. Filled pauses were restricted to *um* and *uh*, but are all shown here as *um*. I have segmented out essentially all non-stop silence as unfilled pause, with the shortest actually measured being only 8 milliseconds long. The reason for not cutting off below some threshold will be discussed in section 3.2 below.

Figure 1, on the following pages, gives the segmentation and duration information I obtained, together with the foot and tone group divisions which the rest of the analysis in this chapter is based on. The tone group boundaries are those from Figure 3 in section 2.2.3 above. Each tone group is printed as a separate paragraph. The foot boundaries are another matter. They are essentially those given in Figure 3 in section 2.2.1 above, but not exactly. I have not included eight of the boundaries marked by subjects A and C only, and have included seven marked only by myself, on the basis of several relistenings to the data segment, paying special attention to those areas of disagreement. This procedure was necessary in order to achieve a consistent and coherent division into feet, which did not emerge from the simple voting method used to arrive at Figure 3 of section 2.2.1.

Also, certain syllables are not included in the analyses in the rest of this chapter, either because they were truncated, that is the speaker started but did not finish a word, or else because they are really outside the foot/tone group structure, being space fillers which serve only to hold the floor. Both these types of syllables are easily identified when listening to the data, and they are indicated in Figure 1 below by being italicized.

```
.62
um (.61) |
------------
.09   .17  .17
 i  / know that (.74) |
------------
.49       .06 .06   .21         .07   .35
um (1.63) on  the / night (.07) be  / fore |
------------
 .10  .12 .05   .21 .12       .38         .10 .15    .24 .27  .04 .09   .27  .15  .09 .18
there was  a  / lot of (.17)  um (.05)   ek  spec / ta  tion  nd  ek  / cite ment on  my (.05)
         .33
      / part (.12) |
------------
.37 .39             .26 .05    .15 .18  .18        .12    .31
and um (1.02) /  af  ter  /  i  went  to  (.24)  to  / sleep (.09) |
------------
.19    .18 .21          .15           .12      .14  .13     .15 .14    .19         .48
my  /  par ents (.05) would (.59) would /  al  ways /  o  pen  /  up  (.84)  um  (.16)
       .12 .19  .06  .07    .30
    /  se  vrel of   the  / gifts |
------------
  .14 .12 .05 .12   .20 .06
/  se vro the im  / por nt |
------------
.27    .23 .08 .12    .43        .29
the /  all the su  / prise (.09) / things (.14) |
------------
.19   .30 .14    .24    .17   .19  .29
and / so  on  / christ mas / mor ning |
------------
.09  .10
you know |
------------
  .13 .14   .18    .22
/  i would wake  / up |
------------
.26      .13   .17 .06 .06 .07 .08    .14 .11 .26
and (.53) go / run ing in  to  the / liv ing room |
------------
.11     .09  .14         .11          .26    .10     .15  .15          .14  .17
and  /  it  would (.04)  be  (.10) / filled with  /  all these (.05) / won drous (.43)
        .31
     / things (.96) |
------------
.68       .88
um (.82) and (1.09) |
------------
.93
then (.08) |
------------
 .13   .16    .20    .22     .15  .17          .07    .21      .13  .27
there would / be  / some / pre  sents (.04) that / werent /  o  pened (.67) |
  .09 .08 .08 .12       .06   .14 .17   .07  .15   .15  .06 .08 .11         .21        .04
/ ty  pi  gli the (.14) the / pre sents that had / been un  der the (.05) / tree (.02) be
      .23    .08 .10 .23
    / fore / a   ny  way (.69) |
______________

.31   .12  .06     .11 .15   .21 .05   .28
and / we  would / o   pen / all of / those (.24) |
______________

.35             .19 .06 .15 .07 .22            .24  .18  .07 .16 .13    .28 .21       .20       .12
and (.36) / i   re  mem ber just (.08) / bing just a   v   a / ve  ry (.44) av (.16) a
      .15 .12    .23 .14    .20 .25         .15 .14         .39
    / re  ly  /  su  per /  ha  py  (.40) / kin of  (.22) / time (.90) |
______________

.72 .09 .16      .26          .41               .34     .08 .19 .07 .10 .08 .28 .06
and we  would / leave (.11) / things (1.15) / spread out all o   ver the k   the
      .14 .12 .15    .61
    / liv ing room / floor |
______________

.52           .21   .08 .08  .33
and (.78) / things in a   / mess (.15) |
______________

  .48 .17 .12 .11  .10   .34         .06 .22 .09   .27  .16    .17     .18    .26
/ for um  you know at / least (.02) the en  at / least all / through that / day (.04) |
______________

  .11  .30    .09 .16   .30
/ some times / se vrel / days (1.11) |
______________

.48   .16  .35   .11  .23   .08 .22
and / that was / kind of / spe cial (.12) |
______________

.06   .11 .09 .08   .28           .25     .11
we / din ave to / clean (.04) / things / up (.82) |
______________

.24 .09 .13       .23
and it  was (.14) / good (3.16) |
______________

.08   .33
the / most (2.35) |
______________

.10   .15  .11    .18  .12     .29    .25  .13        .12 .07   .24 .17   .17 .19 .16
the / thing i / think that / comes / clos est (.04) to  a / fam ly / ri  tu  al (.54) |
______________

.10  .17   .32 .13
in / my / fam ly (.39) |
______________

  .45        .09 .22   .12 .30  .39 .57         .20    .30   .19 .22
/ was (.06)  v   uh / vi  zits to  um (.41) my / grand par ents (.27) |
______________

.09   .38         .43         .14  .06 .09   .36   .05   .36
who / lived (.17) um (.68) / four hun red / miles a / way (.26) |
______________

.32 .59       .11   .16 .11 .20
in  um (.27) new / mek si  ko (1.18) |
.59           .28      .18 .11 .09    .13 .09     .37  .09    .07 .11 .27      .11 .11 .25
and (.20) / that's /  re  ly  the / on  ly  / trips we  /  e  ver took /  a  ny  where (.30) |

.06  .10   .17            .15 .19 .13    .23  .18  .12    .21  .27
 i   mean  my  (.97) /  fa  mli did / not  take  va  /  ca  tions (.07) |

.07 .11  .15     .17     .13  .21               .11  .14 .07    .10  .23    .29
but we  would / go  /  vi   sit  (1.11) / gran ma   n  / gran  pa  / reed (.71) |

.81               .26             .15                .26   .05    .24  .10     .26
um (1.43) / two (.01) / three (.02) / times  a   / year  i  / guess (1.23) |

.61
so (.73) |

  .12   .20  .25              .18   .24 .25              .12  .15           .35
/ there was that (.31) / eight ow  er  (1.02) /  au   to  (.04) / trip (.33) |

.13    .20   .10    .35           .11 .15     .32
 i  / knew  the / road (.04) /  ve  ry  / well (.76) |

.51   .19   .17   .41  .29  .11   .18  .14 .47
and  um / they / lived in  new / mek  si  co |

.34            .15 .29        .15 .09 .09  .07   .19 .20  .09 .11 .13          .13  .26
in (.15) / ve  ry (.18)  ki  na sort of / se  mi / a  re  a (.08) /  a   rid (.14)
         .17  .33
    / coun  try (.42) |

  .34  .20
/ san  dy (1.04) |

.79          .09  .08 .08   .07 .16    .23  .09   .18 .09    .35     .26
and (.68)  so   it  za  / ve  ry  / dif rent / kin of  / world / there (.64) |

.20    .41            .11  .20    .18  .16  .16   .16  .09  .26
and / they (.14) / al  ways / trea ded me / won der  fly (.29) |

.27    .22   .14  .12    .19    .26  .06  .07   .25  .46
my / grand fa  ther / work / tout in  the / oil fields (.04) |

.15 .25 .19    .23  .16    .14  .13           .31   .20
he  was  a  / var ious / kin  of (.05) / sales man (.15) |

.10    .09  .13    .37
at / diff rent / times (.06) |

.29    .13   .14     .19  .13    .16  .23
and / he  would / take me / with him (.41) |

 .09  .09   .25
when he / went (.56) |
 .18   .47
and /  so  (.21) |
__________
  .11   .16     .25 .24
/ we  would / tra  vel (.46) |
__________
  .16 .05  .17
/  i   do  know (.44) |
__________
  .05 .09  .11      .32  .04    .21  .07    .16   .15
/ two hun dred / miles  a   / day  or  / some thing |
__________
 .14   .14            .17 .10  .04     .31
 he  would (.40) / tra  vel   a   / round |
__________
  .18 .08    .20
/ this and / that (2.47) |
```

Figure 1. Duration of all syllables and pauses in data segment

A different view of the overall duration picture is given graphically in Figures 2 - 5 below, which are histograms of the durations of syllables and pauses.[1] Figure 2 covers all the syllables in the data segment. Note that in this figure only each * stands for two syllables.

```
  Duration
 .00  |
 .05  |*************************
 .10  |**********************************************************
 .15  |*******************************************************
 .20  |**********************************
 .25  |***************************
 .30  |***************
 .35  |**********
 .40  |***
 .45  |***
 .50  |**
 .55  |
 .60  |**
 .65  |
 .70  |*
 .75  |
 .80  |*
 .85  |
 .90  |*
```

Figure 2. Histogram of durations of all syllables. Each * indicates 2.

[1] In these and subsequent histograms, all durations have been rounded to the nearest multiple of .05 seconds. Thus the line in the histogram labeled e.g. .25 covers durations from .225 seconds through .275 seconds.

Figure 3 shows the durations of all pauses, and Figures 4 and 5 break these down into filled and unfilled pauses.

```
 Duration
0.00 |****
0.05 |*****************
0.10 |*********
0.15 |**********
0.20 |*****
0.25 |****
0.30 |***
0.35 |**
0.40 |******
0.45 |****
0.50 |
0.55 |***
0.60 |**
0.65 |***
0.70 |***
0.75 |**
0.80 |**
0.85 |*
0.90 |*
0.95 |**
1.00 |**
1.05 |*
1.10 |****
1.15 |*
1.20 |*
1.25 |**
1.30 |*
1.35 |
1.40 |*
1.45 |
1.50 |**
1.55 |
.
.
.
2.05 |
2.10 |*
2.15 |
2.20 |
2.25 |*
2.30 |
2.35 |*
2.40 |
2.45 |*
2.50 |**
```

Figure 3. Histogram of all pauses

```
  Duration
.00  |
.05  |
.10  |
.15  |*
.20  |*
.25  |
.30  |
.35  |
.40  |**
.45  |*
.50  |**
.55  |*
.60  |**
.65  |
.70  |*
```

Figure 4. Histogram of filled pauses

```
  Duration
0.00  |****
0.05  |******************
0.10  |*********
0.15  |************
0.20  |****
0.25  |*****
0.30  |***
0.35  |**
0.40  |******
0.45  |****
0.50  |
0.55  |***
0.60  |**
0.65  |**
0.70  |****
0.75  |***
0.80  |***
0.85  |*
0.90  |*
0.95  |**
1.00  |**
1.05  |*
1.10  |***
1.15  |*
1.20  |*
1.25  |*
1.30  |
1.35  |
1.40  |
1.45  |*
1.50  |
1.55  |
1.60  |
1.65  |*
1.70  |
.
.
.
2.30  |
2.35  |*
2.40  |
2.45  |*
2.50  |*
```

Figure 5. Histogram of unfilled pauses

3.2 Moments of the syllable durations grouped by situation

The structure of feet and tone groups which we have imposed on the data lead to a classification of the syllables according to where they are situated with respect to that structure. After discarding the defective syllables mentioned above, there are 403 syllables in the data segment. Of these, 333 are within feet and 70 are not. Of the syllables not within feet, 45 are upbeat syllables, that is, they occur at the beginning of tone groups, before the first foot. The remaining 25 are what I call *post-pausal* syllables. These are essentially syllables which have fallen through a crack in my transcription system. The directions call for the use of a / to indicate the *beginning* of feet, but do not specify a way to mark the *end* of feet. I did things this way because I found it more natural to listen to data and make single marks as I went along, without having to stop the tape very often to catch up. But this approach leads to the necessity of recognizing this category of post-pausal syllables, which are not really part of any foot. Consider the following excerpt from the data segment:

1) | *there was a / lot of* (.17) $\overset{.38}{\textit{um}}$ (.05) *ek spec / ta tion ...* |

The first three syllables are upbeats, and thus not part of any foot. But what about the first two syllables of the word *expectation*. Over half a second separates them from the previous syllable *of*. If we are to stick to our definition of the foot as a rhythmic whole, structured by the relations between its component syllables, we cannot include the pause and the *ek spec* syllables. The foot then is composed of the syllables *lot of*, and the syllables *ek spec* are not part of any foot. They are in fact very similar to upbeats, being heard as leading up to the subsequent foot boundary. In fact Halliday treats upbeats and post-pausal syllables alike, by using what he calls a *silent ictus*, a sort of silent place holder for the salient syllable of a foot. It occurs as needed at the beginning of tone groups with upbeats, which are then incorporated into a foot. Pauses are handled similarly. Thus he would notate example (1) as follows, using a small wedge ($_\wedge$) for the silent ictus, and double slashes (//) for tone group boundaries:

1') // $_\wedge$ *there was a / lot of /* $_\wedge$ *ek spec / ta tion ...* //

As a result all syllables are part of feet for Halliday. In my opinion it is useful to distinguish the upbeats and post-pausal syllables from those properly within feet.

The three groupings of syllables established so far are distinguished by how a series of syllables begin. It will turn out to be valuable to discriminate also on the basis of what causes us to consider a series to be at an end. The obvious, and in some sense unmarked, case is what I will call the *clean* foot, one which both begins and ends with a foot boundary. But a foot may end because of a pause, as above, or because of the end of a tone group. These I call *pause-end* and *tgb-end* feet respectively. Similarly a series of upbeat or post-pausal syllables may be ended by a foot or tone group boundary, or by a pause. To complete the inventory of the possible situations of a syllable, there is the size of the series within which a syllable occurs, and its position within that series. So for example, in example (1), the syllable *was* is the second of three in a clean upbeat series, and the syllable *of* is the second of two in a pause-end foot. Tables 1 through 5 below show how the duration data looks if we average across all the syllables in the data segment, grouped by the situational categories just established. Not quite all the categories are separated out - upbeats and post-pausal syllables are grouped together, and only distinguished with respect to endings as to clean versus pause-end or tgb-end, to cut down on the number of tables, and because the population in these categories is small in any case. Each row of the tables is for one size group. The first column gives the number of groups of that size, and the subsequent columns give the mean and variance of the durations of syllables in first, second, etc. position in groups of that size.

In considering the tables which follow, it is important to remember that it is the duration of syllables which is given, not individual segments. Some of the variance in the tables therefore comes from the intrinsic difference in the time it takes to articulate syllables with differing numbers of segments. Accounting for this factor would have required a much more detailed phonetic analysis of the data than was practical, but I hope that as the sophistication of computational systems to support such analysis on a large scale increases it will be accounted for, and I trust that the results will corroborate those presented here. I think the size of the sample under consideration here is large enough to balance out this intrinsic variation, and that therefore the results are valid as far as they go.

		Syllable position					
Foot size	n	1	2	3	4	5	6
1	12	.227 .003					
2	41	.186 .006	.136 .004				
3	18	.180 .007	.128 .003	.115 .003			
4	4	.160 .003	.187 .004	.054 .000	.107 .001		
5	1	.166 0.0	.055 0.0	.058 0.0	.067 0.0	.084 0.0	
6	1	.336 0.0	.081 0.0	.193 0.0	.072 0.0	.099 0.0	.078 0.0

Table 1. Moments of durations of syllables in clean feet

The first three rows of Table 1 are the more interesting, as there are not enough feet longer than three syllables to make a good sample for the last three rows. The durations for the clean msfs, trochees and dactyls show a clear pattern, with the first or salient syllable being considerably longer than the others.

		Syllable position				
Foot size	n	1	2	3	4	5
1	36	.304 .007				
2	26	.187 .005	.210 .006			
3	12	.150 .003	.155 .005	.225 .005		
4	3	.169 .009	.098 .002	.083 .000	.134 .002	
5	1	.193 0.0	.064 0.0	.148 0.0	.074 0.0	.222 0.0

Table 2. Moments of durations of syllables in pause-end feet

Again in Table 2, only the first three rows have a sufficiently large population to be significant. Comparing this table to the previous one, we see some similarities and some differences. In particular the last syllable of the foot is considerably longer in the pause-end feet than it is in the clean feet - about 80 milliseconds longer. The msf now really stands out, being as it is both first, and thus salient, and also last, and thus affected by the following pause. These two tables suggest that those two effects are separate and additive.

		Syllable position		
Foot size	n	1	2	3
1	5	.360 .022		
2	3	.180 .000	.166 .014	
3	2	.159 .001	.127 .000	.366 .024

Table 3. Moments of durations of syllables in tgb-end feet

Table 3 looks somewhat like Table 2, but the low sample size and high variance on the diagonal make it difficult to interpret. It seems plausible that the samples which make up this table could be divided into two groups, one of which would pattern like the clean feet, and the other like the pause-end feet, but as there is no independent basis for such a classification, and the number of samples is not large enough to make a significant difference, I have not actually done so.

		Syllable position			
Foot size	n	1	2	3	4
1	26	.112 .003			
2	6	.098 .001	.115 .002		
3	5	.106 .001	.132 .004	.115 .003	
4	1	.150 0.0	.088 0.0	.094 0.0	.074 0.0

Table 4. Moments of durations of syllables not in feet, not immediately pre-pausal

Table 4 covers upbeats and post-pausal syllables whose series end cleanly. There does not seem to be much to chose from here - the grand mean of all 57 syllables is .111, with a variance of only .002.

		Syllable position		
Foot size	n	1	2	3
1	4	.230 .013		
2	3	.106 .001	.123 .001	
3	1	.062 0.0	.098 0.0	.172 0.0

Table 5. Moments of durations of syllables not in feet, immediately pre-pausal or pre-tgb

Table 5 completes the picture, covering upbeats and post-pausal syllables whose series end with a pause or tone group boundary (only one case of the latter). The sample sizes are low, but what there is fits the previous patterns: the last syllable in the series is longer than the corresponding syllable in Table 4 which does not precede a pause.

Some indication of the plausibility of this assertion is given by the pattern of the histograms given in Figures 5a and 5b. Although the sample size is really too small, if you factor in the difference attributable to the msfs being foot-initial as well, and ignore the two extremely long outliers, you can see the two groups, one about 100 milliseconds longer than the other.

```
Duration
.00  |
.05  |
.10  |
.15  |
.20  |*
.25  |
.30  |**
.35  |*
.40  |
.45  |
.50  |
.55  |
.60  |*
```

Figure 5a. Histogram of pre-tgb msf durations

```
Duration
.00  |
.05  |*
.10  |
.15  |*
.20  |
.25  |*
.30  |*
.35  |
.40  |
.45  |*
```

Figure 5b. Histogram of non-msf foot-final pre-tgb syllable durations

3.3 A higher level of structure over the situational groups

A pattern has emerged from the duration data presented in the previous section. It suggests that there is a simple partition of the data which will capture a significant amount of the variation. In this section this partition is first introduced on a relatively informal basis, and then supported by some statistical tests.

3.3.1 Informal presentation

The pattern which I think emerges from the duration data is essentially captured by the positing of two binary features, with the features being ±FOOTINITIAL and (immediately) ±PREPAUSAL. This implies the additional claim that there is no significant difference between syllables within feet and syllables outside the foot structure, except insofar as the feature +FOOTINITIAL implies being within a foot. If we ignore the tgb-end series for the time being, and concentrate on one, two, and three syllable feet[2] this leads to the following organization of the situational groups, where the layout is parallel to the tables in the previous section, and the capital letters denote group membership, with the primed letters are claimed to be not distinguished from the non-primed despite not being part of feet:

		Syllable position		
		1	2	3
Clean feet of length	1	A		
	2	A	B	
	3	A	B	B
Pause-end feet of length	1	C		
	2	A	D	
	3	A	B	D
Clean non-feet of length	1	B′		
	2	B′	B′	
	3	B′	B′	B′
Pause-end non-feet of length	1	D′		
	2	B′	D′	
	3	B′	B′	D′

PREPAUSAL	FOOTINITIAL +	–
–	A	B
+	C	D

Figure 1. Grouping of syllables by features

If we compare the grouping proposed in Figure 1 with the duration means in the previous section, we see that roughly speaking an unmarked syllable is around 100 milliseconds long, and that +FOOTINITIAL and +PREPAUSAL are each worth about an additional 100 milliseconds[3] which collapses the four-way distinction into three durational groups: B, which is 100 milliseconds long, A and D, which are each 200 milliseconds long, and C, which is 300 milliseconds long.

[2] In fact, I have collapsed the feet of more than three syllables into the same rows as the three syllable feet, with the first syllable going to the first column, the last to the last, and the two, three, or four intermediate syllables to the middle.

[3] These figures will be refined in the next section

The next set of histograms supports this division. The first is for the six groups combined, and the other six are for each group separately.

```
 Duration
.00  |
.05  |***********************
.10  |******************************************************
.15  |**************************************************
.20  |******************************
.25  |************************
.30  |***********
.35  |********
.40  |***
.45  |**
.50  |*
```

Figure 2a. Histogram of 6 groups of syllables combined. *Each * represents 2 instances.*

```
     A: +FootInitial -PrePausal                 B: -FootInitial -PrePausal                      B′: -FootInitial -PrePausal
     in feet, salient                           in feet, non-salient                            not in feet
.00  |                                          |                                               |
.05  |***                                       |***************************                    |**************
.10  |**********************                    |**********************************************  |******************************
.15  |****************************************  |*****************************                  |***********
.20  |*************************                 |*************                                  |****
.25  |*********************                     |*****                                          |**
.30  |********                                  |***                                            |
.35  |*****                                     |*                                              |
.40  |*                                         |                                               |
     N=124 mean=.185 var=.005                   N=121 mean=.125 var=.004                        N=61 mean=.111 var=.002
```

Figure 2b. Histograms of –PrePausal groups, A, B, and B′.

Figure 2b graphically points out the shift in duration between group A, which is the +FootInitial, or salient, syllables, and group B, the –FootInitial, or non-salient, syllables, and also the lack of difference between group B and group B′, the –PrePausal syllables which are not within feet.

```
     C: +FOOTINITIAL +PREPAUSAL    D: -FOOTINITIAL +PREPAUSAL   D′: -FOOTINITIAL +PREPAUSAL
     in feet (+PREPAUSAL msfs)     in feet                      not in feet
.00  |                             |                            |
.05  |                             |                            |
.10  |*                            |***                         |*
.15  |*                            |*************               |****
.20  |****                         |***********                 |
.25  |*******                      |***********                 |
.30  |********                     |*                           |*
.35  |********                     |*                           |*
.40  |****                         |*                           |
.45  |**                           |*                           |
.50  |*                            |                            |
     N=36 mean=.304 var=.007       N=42 mean=.209 var=.005      N=7 mean=.195 var=.009
```

Figure 2c. Histograms of +PREPAUSAL groups, C, D, and D′.

Figure 2c parallels Figure 2b for the +PREPAUSAL syllables, and the same pattern appears: D is shifted from C, but D and D′ are similar.

3.3.2 Statistical support

3.3.2.1 Linear regression

There are a number of different ways in which more formal statistical support for the proposed feature analysis might be sought. Since I am proposing not only a feature analysis, but one in which the features contribute linearly to the total duration of a given syllable, one possible statistical test is a linear regression, to see how much of the variance in the syllable durations is accounted for by the proposed feature analysis. We will start with a simple case, considering only short, clean feet, that is, msfs, trochees, and dactyls which end with a foot boundary. Table 1 below shows the results of the regression, where duration is the dependent variable and FOOTINITIAL (taking value 0 or 1) is the independent variable.

	B	SE	EtaSq	F	df	p
FOOTINITIAL	.062	.011	.175	30.874	1.000	.000
Constant	.129	.008	1.559	275.739	1.000	0.000
Regression		.379	.175	30.874	1.000	.000
Residual		.068	.825		146.000	

Table 1. Regression of duration of short clean feet as a function of FOOTINITIAL

Reading these regression tables is not as hard as it might seem. The first column gives the coefficients for a linear equation expressing the dependent variable as a function of the independent variable(s). In this case, the equation is *duration* = .062 · FOOTINITIAL + .129. In other words, the best linear prediction of the duration of these syllables is that the –FOOTINITIAL syllables are .129 seconds long, and the +FOOTINITIAL syllables are .062 seconds longer. The other columns of the table tell how good an approximation this equation actually is to the data. In particular the column labeled *EtaSq*, for eta squared, tells what percentage of the variance is accounted for by the independent variable(s), and by the whole equation (in the row labelled *Regression*). In this case, as we have only one independent variable, these are the same, namely 17.5%, leaving 82.5% of the variance as residual, or unaccounted for. The other column of interest is the one headed *p*, which gives the significance level of the eta squared values. As all the values in these tables are rounded to the nearest .001, we see that in this case the 17.5% figure is significant to .the .0005 level or better.[4]

Accounting for only 17.5% of the variance may not seem like much, but we must bear in mind that each of our original situational groupings had a non-trivial amount of variance, which we cannot hope to account for. We can get some idea of how well we are doing by taking the situational grouping as the independent variable, and seeing how well *that* does. Table 2 shows the results.

[4] Values of exactly 0 can be distinguished from values somewhere between 0 and .0005 because the former is printed as 0.000 while the latter appear as .000

	B	SE	EtaSq	F	df	p
One1	.112	.025	.109	19.471	1.000	.000
Two1	.071	.019	.077	13.655	1.000	.000
Two2	.020	.019	.006	1.103	1.000	.295
Three1	.064	.023	.045	8.052	1.000	.005
Three2	.013	.023	.002	.335	1.000	.564
Constant	.115	.016	.291	51.827	1.000	0.000
Regression		.183	.204	7.262	5.000	.000
Residual		.068	.796		142.000	

Table 2. Regression of duration of syllables in short clean feet as a function of situational grouping

The five independent variables, which are given values of 0 or 1, encode the situational group membership of the syllables in a mutually exclusive way. *One1* covers the msf syllables, *Two1* the first syllables of trochees, *Two2* the second syllables of trochees, and so on. There is no *Three3* as that group is distinguished by 0 values for the five other variables.

A number of interesting things emerge from this table. First of all, this complete partition into situational groupings only accounts for 20.4% of the total variance, suggesting that the 17.5% accounted for by the FootInitial feature is in fact quite a respectable performance. We also see that the distinction between group Three3, represented by the constant term, and both Two2 and Three2 is small, 20 and 13 milliseconds respectively, does not make a substantial contribution to the variance, .6% and .2% respectively, and in any case that contribution is not significant, p=.295 and p=.564 respectively. These three groups are the constituents of the feature group B, and these figures support their being grouped together. There is also some suggestion that the msfs are longer than the initial syllables of the trochees and dactyls, with an incremental contribution of 112 ms. as opposed to 71 and 64. It turns out that this distinction is not significant, but I will leave off discussion of that until the end of this section.

The next step is to expand the scope of the regression to all the syllables (although still excluding the 11 pre-tgb syllables) and both features. The results of this are given in Table 3.

	B	SE	EtaSq	F	df	p
FootInitial	.072	.007	.169	110.373	1.000	0.000
PrePausal	.100	.008	.229	149.317	1.000	0.000
Constant	.117	.005	.936	611.134	1.000	0.000
Regression		.770	.404	131.993	2.000	0.000
Residual		.067	.596		389.000	

Table 3. Regression of duration of all –pre-tgb syllables as a function of FootInitial and PrePausal

The increase in sample size and the addition of the additional feature have increased the percentage of variance accounted for to 40.4%, which is quite good given the variances within the situational groups. If we also consider that on the order of 40% of the variance is attributable to segmental effects (Mark Liberman, personal communication), this is about is good as we could hope for, as there is bound to be some random fluctuation. The equation we get agrees well with what we would expect based on the means in the previous section: *duration* = .072 · FootInitial + .100 · PrePausal + .117. This gives durations of 117 ms. for groups B and B′, 199 ms. for group A, 217 ms. for groups D and D′, and 299 ms. for group C.

3.3.2.2 Analysis of variance, Newman-Keuls

Another method of validating the distinctions I have suggested is by doing an analysis of variance with respect to various proposed groupings, to determine how the within-group variance compares to the between-group variance, which in turn gives a measure of the significance of the grouping. First I will give some simple analysis of variance results for some of the simple groupings, and then a couple of larger scale analyses.

If we consider once again the clean trochees and dactyls, we find support for the obvious effect of position within the foot:

	Moment			
Position	N	Mean	Variance	
1	41.000	.186	.006	
2	41.000	.136	.004	p<.002

Table 1a. Moments of duration of trochee syllables by position

	Moment			
Position	N	Mean	Variance	
1	18.000	.180	.007	
2	18.000	.128	.003	
3	18.000	.115	.003	p<.012

Table 1b. Moments of duration of dactyl syllables by position

The p values are for the significance of the effect of position - presumably somewhat less for the dactyls as there is little or no difference between positions 2 and 3.

Tables 1a and 1b gives the significance of the difference of the means. The mean of the differences is also significant:

Moment			
N	Mean	Variance	
41.000	.051	.015	p<.01

Table 2a. Moment of difference in duration of first and second syllables of clean trochees

Moment			
N	Mean	Variance	
18.000	.051	.005	p<.007

Table 2b. Moment of difference in duration of first and second syllables of clean dactyls

Here the p values measure the significance of the difference between the mean of the differences and 0, which in turn is a measure of the significance of the mean of the differences itself.

For the dactyls what we would really like is some analysis of the pairwise relationships for all possible pairs of positions. The Newman-Keuls test allows us to do that. As we will use this test several more times, I will go through the derivation this time, where the number of moments is small. We start with an ordered set of means, such as the one in Table 1b. From this we compute an array of pairwise differences:

Position		
2	1	Position
.013	.064	3
	.051	2

Table 3a. Pairwise differences of means of duration of dactyllic syllables by position

To obtain F values, we then divide each element of this array by the square-root of the quotient of the mean-squared error from the analysis of variance of the means, which gave us the p value in Table 1b, and the harmonic mean of the number of samples in each cell of the ordered set of means: $\sqrt{MS_E/n_h} = \sqrt{.004/18} = .015$:

Position		
2	1	Position
.848	4.156	3
	3.309	2

Table 3b. F values derived from Table 3a

We then look up these values in a table of the studentized range statistic, using the row determined by the degrees of freedom for the error term in the analysis of variance, in this case 51, and the column determined by the distance separating each pair of means in the original set, in this case 3 for the upper right hand element and 2 for the other two, where we discover that the underlined cells in table 3b are significant to the .05 level, and the other cell is not. From this we conclude that in clean dactyls, the first syllable is significantly different from both the second and the third, but that the second is not significantly different from the third, which is just what we thought.

The Newman-Keuls test can be applied to any number of means, and I have applied it to the means of the situational groups for the clean and pause-end feet. The ordered set of moments is as follows, where the group labels are similar to those use above in section 3.3.2.1, Table 2, with the prefixing of a 'P' for the pause-end groups, so that e.g. *Two2* covers the second syllables of clean trochees and *PTh2* covers the second syllables of pause-end dactyls.

	Moment		
Group	N	Mean	Variance
Three3	24.000	.111	.002
Three2	33.000	.118	.004
PTh2	21.000	.128	.004
Two2	41.000	.136	.004
PTh1	16.000	.156	.004
Three1	24.000	.182	.007
Two1	41.000	.186	.006
PTw1	26.000	.187	.005
PTh3	16.000	.207	.005
PTw2	26.000	.210	.006
One1	12.000	.227	.003
PO1	36.000	.304	.007

Table 4a. Ordered moments of durations of syllables from clean and pause-end feet by situational group

This gives the following table of pairwise differences:

Group											
Three2	PTh2	Two2	PTh1	Three1	Two1	PTw1	PTh3	PTw2	One1	PO1	Group
.007	.017	.024	.045	.071	.075	.076	.096	.099	.116	.193	Three3
	.010	.017	.038	.064	.068	.069	.089	.091	.109	.185	Three2
		.008	.028	.055	.059	.059	.080	.082	.099	.176	PTh2
			.020	.047	.051	.051	.072	.074	.092	.168	Two2
				.026	.030	.031	.051	.054	.071	.148	PTh1
					.004	.004	.025	.027	.045	.121	Three1
						.000	.021	.023	.041	.117	Two1
							.021	.023	.040	.117	PTw1
								.002	.020	.096	PTh3
									.017	.094	PTw2
										.076	One1

Table 4b. Pairwise differences of means from Table 4a

The analysis of variance of Table 4a and the harmonic mean of the Ns gives us the constant for the division, which in turn gives as the array of F values:

	SumSq	df	MS	F	p
Gnd-mean	10.145	1.000	10.145	2067.132	0.000
Factor1	1.002	11.000	.091	18.562	.000
Error	1.492	304.000	.005		

Table 4c. Analysis of variance of Table 4a

$\sqrt{MS_E/\bar{n}_h} = \sqrt{.005/22.9} = .0146$

Figure 1. Calculation of constant divisor

Group											
Three2	PTh2	Two2	PTh1	Three1	Two1	PTw1	PTh3	PTw2	One1	PO1	Group
.488	1.142	1.671	3.072	4.870	5.148	5.177	6.582	6.743	7.936	13.165	Three3
	.654	1.183	2.583	4.382	4.660	4.689	6.093	6.255	7.448	12.676	Three2
		.528	1.929	3.728	4.006	4.035	5.439	5.601	6.794	12.022	PTh2
			1.401	3.199	3.477	3.506	4.911	5.072	6.265	11.494	Two2
				1.798	2.077	2.106	3.510	3.672	4.864	10.093	PTh1
					.278	.307	1.712	1.873	3.066	8.295	Three1
						.029	1.433	1.595	2.788	8.016	Two1
							1.404	1.566	2.759	7.988	PTw1
								.162	1.354	6.583	PTh3
									1.193	6.422	PTw2
										5.229	One1

Table 4d. F values derived from quotient of Table 4b and the constant from Figure 1

Finally, looking these values up in the studentized range statistic table gives the following pattern of differences significant to the .01 level, where a + indicates a significant difference, and - indicates no significant difference:

Group Three2	PTh2	Two2	PTh1	Three1	Two1	PTw1	PTh3	PTw2	One1	PO1	Group
-	-	-	-	+	+	+	+	+	+	+	Three3
	-	-	-	-	-	-	+	+	+	+	Three2
		-	-	-	-	-	+	+	+	+	PTh2
			-	-	-	-	+	+	+	+	Two2
				-	-	-	-	-	-	+	PTh1
					-	-	-	-	-	+	Three1
						-	-	-	-	+	Two1
							-	-	-	+	PTw1
								-	-	+	PTh3
									-	+	PTw2
										+	One1

Table 4e. Pairwise differences significant to the .01 level

This pattern of significantly distinct pairs can be interpreted to associate the groups in the following pattern, where each underlined subset represents a group within which no significant difference can be found:

PO1 One1 PTw2 PTh3 PTw1 Two1 Three1 PTh1 Two2 PTh2 Three2 Three3

Figure 2. Significant associations of situational groups

This is not perfect, but close to it. We have PO1, which is group C, clearly distinct, and the next 7, which together are A and D, all grouped together as we would predict. Unfortunately, group B does not cohere perfectly, shading up into the lower components of A, but in general these results add to the plausibility of the proposed feature analysis.

3.3.3 Whole foot durations

Another way of smoothing out some of the variance is to look at the total duration of feet consisting of a given number of syllables, and the next two tables give this data for clean and pause-end feet:

	Foot Size					
Moment	1	2	3	4	5	6
N	12.000	41.000	18.000	4.000	1.000	1.000
Mean	.227	.322	.424	.509	.430	.859
Var	.003	.005	.016	.009	0.000	0.000

Table 1a. Moments of total durations of clean feet by foot size

	Foot Size					
Moment	1	2	3	4	5	6
N	36.000	26.000	12.000	3.000	1.000	0.000
Mean	.304	.397	.529	.483	.700	
Var	.007	.012	.019	.032	0.000	

Table 1b. Moments of total durations of pause-end feet by foot size

As the graphical version of this data shows, the curves are quite linear in the number of syllables, as long as the sample size holds up:

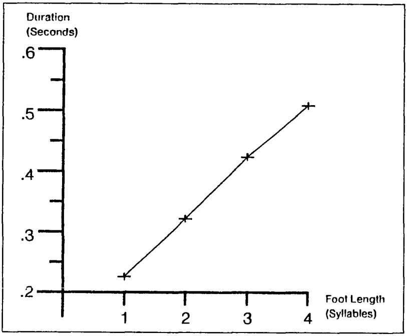


Figure 1a. Graph of means from Table 1a

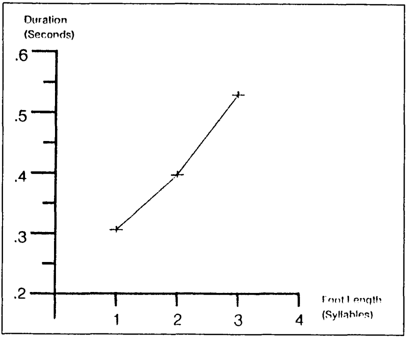


Figure 1b. Graph of means from Table 1b

Further support for the linearity of the relation is provided by the following regression of total duration as a function of length in syllables:

	B	SE	EtaSq	F	df	p
# of syllables	.097	.013	.421	53.017	1.000	0.000
Constant	.130	.031	.142	17.917	1.000	0.000
Regression		.636	.421	53.017	1.000	0.000
Residual		.087	.579		73.000	

Table 2. Regression of duration of clean feet 4 or fewer syllables in length as a function of length in syllables

The incremental value of 97 ms. is in the same ballpark as the other values we have seen for –FootInitial syllables.

3.3.4 *Leftovers*

Finally there are a few more statistical results in an attempt to forestall comments of the form "But why didn't you check the relation between ... and ... ?"

First, as promised above, several indications that there is no significant difference between the duration of clean msfs and the first syllable of clean trochees and dactyls. We have the following means and analysis of variance:

	Moment		
Size	N	Mean	Variance
3	18.000	.180	.007
2	41.000	.186	.006
1	12.000	.227	.003

Table 1. Ordered moments of duration of foot-initial syllables from clean feet by size of foot

	SumSq	df	MS	F	p
Gnd-mean	2.607	1.000	2.607	461.803	0.000
Factor1	.019	2.000	.009	1.670	.196
Error	.384	68.000	.006		

Table 2. Analysis of variance of Table 1

The p value suggests not much is going on, and a Newman-Keuls test supports that conclusion, with the following array of F values, none of which are significant at the .05 level:

Position		
2	1	Position
.384	2.710	3
	2.326	2

Table 3. F values for Newman-Keuls test on Table 1

Further lack of support for a distinction here comes from the following regression, which adds an independent variable coding for msfs:

	B	SE	EtaSq	F	df	p
FootInitial	.055	.012	.123	22.269	1.000	.000
msf	.043	.021	.022	4.005	1.000	.047
Constant	.129	.008	1.559	281.413	1.000	0.000
Regression		.285	.197	17.757	2.000	.000
Residual		.068	.803		145.000	

Table 4. Regression of duration of syllables in short clean feet as a function of FootInitial and msf

The new feature only accounts for 2.2% of the variance, and that only with a p value of .047, which is much less good than the <.0005 of the others.

Finally, if we do not restrict our sample by throwing out the pre-tgb syllables, as we have done above, but rather consider all 403 syllables, and include in the analysis an additional independent variable coding pre-tgb status, we do just as well, in fact slightly better:

	B	SE	EtaSq	F	df	p
FOOTINITIAL	.074	.007	.161	108.920	1.000	.000
PREPAUSAL	.100	.009	.201	136.008	1.000	0.000
pre-tgb	.134	.022	.057	38.834	1.000	0.000
Constant	.116	.005	.820	555.278	1.000	0.000
Regression		.676	.411	92.710	3.000	0.000
Residual		.070	.589		399.000	

Table 5. Regression of duration of all syllables as a function of FOOTINITIAL, PREPAUSAL, and pre-tgb

3.4 Conclusions

What does this mass of statistics mean? It would seem that, based on all these numbers, and the sense of the patterns behind them which I hope has emerged, isochronicity is at best a misleading hypothesis about English. Any tendency to isochronicity would have to be revealed as a non-linearity in the durations of feet of increasing size. But all the numbers so far seem to support rather than disconfirm a linear relation. The simpler analysis in terms of the features FOOTINITIAL and PREPAUSAL seems both more accurate and more plausible - it has long been established that duration is one of the principal determinates of perceived accent, and the lengthening of syllables before a pause also seems to make sense in terms of the dynamics of the speech production process - if you don't know exactly what to say next, you lengthen the last thing you *do* have to say in an attempt to bridge the gap. I will pursue this latter idea further in the next chapter.

I think the success of the isochronicity hypothesis anecdotally can in part be accounted for by the fact that so many feet are three or fewer syllables long - around 90% in my data, breaking down approximately 30%-43%-17% for msfs, dactyls, and trochees respectively. Over this range the most obvious fact is that dactyls are only twice as long as msfs, which seems to be a non-linearity. But once we descend below the level of the foot to the syllables, the linearity becomes apparent, and isochronicity seems to be a misleading way to characterize the evident regularity of foot and syllable duration.

Chapter 4. The relation of stress and foot structure: A production model

Drawing on the results presented in the two previous chapters, in this chapter we address the grammar of foot structure. That is, we are concerned with characterizing the possible division into feet of utterances of a given sentence. In service of this goal, a model of a part of the production process is proposed, called the footmaker, with two principal aims: To bring together and make explicit all the extant proposals relating lexical stress, form class, syntax, and foot structure, and to account for some of the observed variability in foot structure in terms of the temporal aspects of the model.

4.1 The problem and what has been said about it

If all the words in all utterances occurred in citation form, that is, with all and only the stressed syllables appearing as salient, then there would be no problem. But in fact this is not the case, and variation is possible. For instance, in the data segment, the words *over* and *under* occur as non-salient. Some monosyllabic words always appear as non-salient, others are always salient, others still appear in both forms.

The problem then is to determine what the relationship is between the stress patterns of the individual words in a sentence and the possible division into feet of utterances of that sentence, given that it is not a simple one-to-one mapping. We must determine both what aspects of the sentence in addition to the stress patterns of its constituent words are relevant, and how they affect the foot structure of an utterance.

4.1.1 The simple hypothesis

What appears at first to be quite a satisfactory solution to the problem can be expressed in simple terms, and is all one usually finds in introductory texts. We distinguish *content* words from *function* words, or as it is sometimes put, *open class* words from *closed class* words. The first group includes at least nouns, verbs, and adjectives - those form classes which have an effectively unbounded number of members, and which have obvious semantic content. The second group includes at least articles, pronouns, and prepositions - those form classes which have a fairly small, fixed number of members, and whose semantic content is in many cases limited. We then claim that in uttering a sentence just the stressed syllables of the content words are realized as salient, and all other syllables, that is all unstressed syllables and the stressed syllables of function words, are realized as non-salient.

There are a number of ways in which this simple hypothesis is not altogether accurate, many of which have been noted in the past. The balance of this section is devoted to an enumeration of exceptions to this approach. The phenomena discussed here will each be enlarged on and formalized in the specification of the footmaker which follows in section 4.3.

4.1.2 The Rhythm Rule

The *Rhythm Rule* (also know as the *thirteen men* rule) is the name given to the phenomenon exemplified by the following pattern of foot structure:

1a) */ Robin / loves Tenne / see*
b) */ Robin / lives on the / Tennessee / border.*

In the abstract, *Tennessee* is stressed on its last syllable, and in citation form or examples such as (1a), that syllable is salient, in accord with the simple hypothesis. But example (1b) manifests a different structure, where the salience has been retracted to the first syllable. The foot structure predicted by the simple hypothesis -

1c) *? / Robin / lives on the Tenne / see / border*

is possible, but clearly (1b) is preferred. This phenomenon is much more common in speech than might be supposed, and its proper analysis has been of some concern in the field. Some representative samples are given below. In each case the first foot boundary as marked occurs earlier (in the context of the whole phrase as cited) than the simple hypothesis would predict, as can be seen by considering where the foot boundary falls when the underlined portion appears alone.

2a) */ thirteen / men* vs. *thir / teen*
b) */ Artificial In / telligence* vs. *arti / ficial*
c) */ achromatic / lens* vs. *achro / matic*
d) */ good-looking / life guard* vs. *good / looking*
e) */ Golden Gate / bridge* vs. *Golden / Gate*

Subjectively the correct characterization is clear - the pattern called for by the simple hypothesis is awkward. There is a sense of rhythmic clash brought about by the adjunction of the two salient syllables. By far the most complete and satisfying theoretical approach to this problem is that of Liberman and Prince [1977]. They use two formal structures to predict when the Rhythm Rule may apply. One is their solution to the problem of stress in general, which provides for each word a relational *metrical structure*, and the other is a notion of *metrical grid*, against which metrical structures are assembled in the composition of utterances. It is possible to characterize the situations when the Rhythm Rule may apply in terms of particular relationships between the structure and the grid. This approach will be discussed further below, in section 4.3, where we will adopt a version into the footmaker.

4.1.3 The pressure towards alternation

Put simply, other things being equal, speakers of English seem to prefer trochees (two syllable feet) and dactyls (three syllable feet) over msfs on the one hand and longer feet on the other. The distribution of the length of the feet in the data segment bears this out - 88% of the clean feet are trochees or dactyls. This is not just epiphenomenal - there seems to be a significant pressure away from sequences of salient syllables without any non-salient syllables intervening, which is to say away from msfs, and away from sequences of non-salient syllables without any salient syllables intervening, which is to say away from long feet.

This pressure manifests itself in various ways. The Rhythm Rule, mentioned above, is one. A related phenomenon is exemplified by the following excerpt from the data segment:

1) *at / least all / through that / day.*

In the balance of the text, mono-syllabic quantifiers such as *all* nearly always occur as salient, while mono-syllabic prepositions such as *through* nearly always occur as non-salient. Indeed, a version of (1) conforming to that pattern is possible, namely:

1a) *at / least / all through that / day.*

But this does indeed feel awkward in a way that (1) does not. A version of the Rhythm Rule is at work here, which will be discussed below in section 4.3.5.

Disyllabic prepositions and some disyllabic conjunctions sometimes appear as salient, sometimes not. The conditioning for this seems to be mostly the pressure for alternation. Consider the following:

2a) *The / books under the / table are / good.*
2b) *The / newspapers / under the / table are / good.*
2c) *? The / newspapers under the / table are / good.*
2d) *? The / books / under the / table are / good.*

3a) */ Under the / table / crouched a / bear.*
3b) *? Under the / table / crouched a / bear.*

It appears that the presence or absence of a foot boundary before *under* is determined by which situation would give the "best" foot structure - that is, one without msfs ((2a) vs. (2d)) or very long feet ((2b) vs. (2c)) or long upbeats ((3a vs. 3b)). In words with many syllables, secondary stressed syllables may appear as salient under similar conditions, e.g.:

4a) *They pre / fer cooper / ation.*
4b) *They / talked about / cooper / ation.*

Note that this pressure towards trochees and dactyls is consistent with our underlying analysis of foot structure as a rhythmic, relational phenomenon. Sensible alternation is constitutive of such a phenomenon, and without it foot structure would cease to be perceptible at all. Long strings of msfs or non-salient syllables do not provide the necessary alternation, and are therefore avoided when possible.

4.1.4 Interaction with highlighting

Highlighted syllables are always salient. Thus articles and even the unstressed syllables of content words may occur as salient if there is some semantic reason to highlight them, as in the contrastive cases, e.g.:

1a} *I said /* $\overset{\uparrow}{\textit{a}}$ *good / reason not /* $\overset{\uparrow}{\textit{the}}$ *good / reason*

b) *I / said it /* $\overset{\uparrow}{\textit{depressed}}$ *me / not that it /* $\overset{\uparrow}{\textit{impressed}}$ *me.*

There is also an accompanying phenomenon which functions to bleach out the salience of immediately following syllables, as if the highlighting had in some sense raised the ante:

2) *I said /* $\overset{\uparrow}{\textit{a}}$ *reason not /* $\overset{\uparrow}{\textit{the}}$ *reason.*

Despite the fact that *reason* has stress on its first syllable, no foot begins there in (2), apparently because the highlighted articles are so prominent.

4.1.5 Interaction with pauses

All the above discussion has tacitly assumed that utterances consist of strings of clean feet without interruption. The fact that this is not true introduces another potential source of variation. What might in one case be a four syllable foot, and thus marked, and a target for enforced alternation, turns into a three syllable foot and a one syllable upbeat when a pause is introduced - no problem. This and similar phenomena occur frequently in the text, for instance:

1) *... / treated me / wonderfully (.29) my / grandfather ...*

Constituent boundaries have an effect similar to that of pauses, at least in some cases. It has been noted for example that the Rhythm Rule rarely if ever operates across a major constituent boundary.

4.1.6 Other interactions

Content words may appear as non-salient if they are sufficiently lacking in content, that is if they are redundant, more or less predictable from context. This shows up most with respect to transitive verbs, and, at least in a context of actual repetition, can be viewed as the last step before complete disappearance, that is, gapping. In the text we see for instance:

1) */ that's / really the / only / trips we / ever took / anywhere |*
I mean my / family did / not take va / cations.

Here neither *took* nor *take* are salient, the first being essentially predicted by the previous occurrence of its cognate object *trips*, and the later being a repetition. Note also the non-salient occurrence of *mean*. This bleaching of verbs in parenthetical/modal contexts would appear to be a related effect.

A sort of inverse of this occurs with pronouns. Many times they are simply place fillers, required by the grammar but totally predictable. In this case they appear as non-salient. But they may be salient, without being highlighted, just in case the choice of pronoun represents a real semantic choice. In the text this usually occurs with respect to subject pronouns, when a new theme is introduced. For instance compare the two instances of *we* in the following:

2) *(I remember just being a .. super happy kind of time)*
/ we would / leave / things / spread out ...
(and that was kind of special)
we / didn't have to / clean things / up.

There is also presumably some interaction with the pressure towards alternation here - it is not always possible to distinguish just what is going on, but the question of redundancy once again seems relevant.

Compounds, especially short ones, behave differently from similar sequences of words which do not form a compound. One classic example is

3a) */ steel warehouse*
b) */ steel / warehouse,*

where the first, a compound, describes a warehouse for storing steel, while the second, presumably not a compound but a noun phrase consisting of a denominal adjective and a noun, describes a warehouse made of steel. In longer compounds the difference may tend to disappear under the pressure for alternation, as in */ living room / floor* which appears in the text.

This is in fact part of a larger phenomenon. In general, every element in group and clause structure (verb group, noun group, clause) has an information focus. This is the site of principal communicative content, and normally in English is at the right hand end of the unit. Kinetic tones are associated with the focus of the clause, and highlighting of a whole group is accomplished by highlighting the focus. But marked (non-unit-final) focus is possible. In particular, it is the norm for noun-noun compounds, as it is the modifying noun which contributes most to identifying the referent.

Terry Winograd (personal communication) has suggested a figure-ground distinction as the basis for the difference between (3a) and (3b), instead of the above-mentioned grammatical one. The figure is focused with respect to the ground. So in (3a), where the implicit contrast is with other sorts of warehouses, we have something like "In the context of *warehouses* (= ground), one for *steel* (= figure)." In (3b), where the implicit contrast is with other things made of steel, we have something like "In the context of *steel* (= ground), a *warehouse* (= figure) made of it." This seems a more perspicuous distinction than that between noun-noun and adjective-noun structure, although they are clearly related. Consider e.g. */ Stanford game* versus */ Stanford / library,* where the contrast between "In the context of games, the one with Stanford." and "In the context of Stanford, its library." seems to get at what is going on here better than calling *Stanford* a noun in the first case and an adjective in the other.

Be that as it may, we can coherently account for the bleaching which occurs here, as well as the previously discussed bleaching following highlighting, as well as some complex interactions between compounding and the Rhythm Rule, by supposing that marked focus within any domain reduces the potential for salience of the defocused (subsequent) elements in that domain.

And finally, in a related phenomenon, the two parts of a verb-particle construction are realized differently depending on whether the particle is separated from the verb or not, as in the following examples from the text:

4a) */ I would wake / up ...*
b) *we / didn't have to / clean things / up.*

There seems to be a tendency, particularly among monosyllabic verbs, to occur as non-salient when the particle is not moved, as in (4a), but salient if the particle is moved, as in (4b).

All these phenomena need to be accounted for by formal specification within a theoretical framework, and the footmaker as I will specify it below is an attempt to do this. In particular I contend that the process modelling approach is well suited to many aspects of this task, and that many of the above patterns can be elegantly captured thereby.

4.2 Modeling the production process - where the footmaker comes in

When we turn our attention from competence to performance, from an abstract structural domain to a concrete temporal one, several radical changes in approach are required. In particular, the nature of the human brain and the temporal organization of speech become relevant constraints on the types of models which can be proposed. If we are to specify the properties of the footmaker, we must have some sense of the overall structure within which it is situated, and of some of the fundamental constraints on that structure.

4.2.1 The nature of process models: Taking the left to right constraint seriously

It is seductive to suppose that a transformational grammar *is* a production model - one starts with an idea, encodes it in deep structure, transforms it into surface structure, and utters it. Leaving aside the Generative Semantics vs. Extended Standard Theory debate about whether that is even satisfactory as a *competence theory,* it is clear for independent reasons that it cannot be a performance model, because it violates a strong, although admittedly largely subjectively based, constraint; which I call the *left to right* constraint. This constraint reflects the temporal organization of spoken language, and holds that it is an essential feature of a plausible processing model that the processing (for either production or recognition) of the beginning of an utterance must not be critically dependent on the prior or simultaneous processing of the end of the utterance. In simpler terms, that the model must be able to start an utterance without knowing how it will end. The naive transformational model clearly violates this constraint.

The other major constraint we must consider is the brain itself. Considered as a processing device, it is not particularly fast, being incapable of performing more than about ten operations a second of the sort which might reasonably be supposed to be involved in language production, *provided that each operation depends on the previous ones completion.* That is to say, reaction time studies (e.g. [Newell 1973], [Sternberg 1969]) suggest that there is a sort of fundamental unit task in the operation of the higher faculties of the brain, and that it takes the brain about a tenth of a second to perform this task. This limitation is rarely felt, however, since independent operations may proceed in parallel, although there is some possibility of interference, and a limit to how much can be going on at one time.

Given the slow speed of the brain, and the fact that in the production of utterances there are clearly a number of steps which are necessarily ordered one after the other, viz. words within a constituent must be chosen and ordered before they can be said, either a subject is chosen and this determines voice, or vice versa, etc., it would seem that speech as we know it would be impossible. We would expect long pauses between, say, constituents, as the silent parts of the production process happened. But the parallelism of the brain provides a solution. We can suppose that e.g. the actual uttering of one constituent may proceed in parallel with the lexicalization and ordering of the next constituent, and so on, so that at any given moment, different parts of an utterance are at different stages in the process of reaching utterable form.

In computer science, this use of parallelism and decomposition into stages is called *pipelining.* The industrial assembly line is the prototypical example of pipelining. If the production of e.g. a toy takes three steps of equal duration, namely cutting the wood, screwing the pieces together, and painting, then one man can produce a toy every three units of time, while three men, each specialized to a particular task, can produce a toy every unit of time, if one cuts the wood, the next takes the pieces from him and assembles them, and the third paints. At any given time, there are three toys on the assembly line, at various stages of completion. To belabor the obvious, the secret is that each man works continually, not waiting for the results of his efforts to emerge. Note that if there is a delay at any stage, it will propagate from stage to stage, and emerge from the assembly line as a delay in the final product.

We see this in speech, where investigators of pauses have long distinguished between pauses which, in our terms, were introduced at different stages in the assembly line. For instance we can distinguish pauses which occur because of indecision over what participant to identify next, as opposed to those stemming from inability to determine *how* to refer to a particular participant. When the pauses are unfilled, this is not always easy to determine, but as James [1973] has shown, the different sorts of filled pauses diagnose different sorts of delays in the production process - in our terms they identify the stage of the assembly line at which the hold up occurred. Chafe [1979b] also discusses the issue of different sorts of pauses in a way compatible with this approach.

The natural question which then arises is: What are the stages into which the speech production process is divided? There is of course much room for argument here, but as a first cut I think in terms of a five stage model - Framing, Coding, Grammar, Phonology, and Articulation. The footmaker is a sub-stage of Phonology. At this early stage in the investigation of process models, it is important to recognize that such a statement is not to be taken as saying "This is all you need and this is how it's organized", but rather as suggesting that these at least are recognizable sub-tasks of the production process, which seem to meet the criteria set forth above for being stages in an assembly line, namely they are relatively independent one from the other, and they are ordered in time. The assumption of independence and ordering is in part a methodological one. At the beginning of such an investigation, it is a plausible simplifying assumption. Examples which suggest an influence on an earlier stage by a later one are not disasters, merely indications of where attention must be paid after the framework has been laid out.

Underlying this proposal is the assumption that after the Framing stage the unit of production we are considering is for the most part the phrase or group. That the items in the assembly line must be smaller than the sentence follows from the left-to-right constraint. Their identification with e.g. noun phrases, prepositional phrases, and verb groups is attractive but certainly not firmly established. We will note evidence both pro and con below.

What follows in the next three sections is a sketch of the five stages of the proposed model, showing where the footmaker is situated therein.

4.2.2 What comes before: Framing, coding and grammar

In [Thompson 1977] I presented a partial model of two aspects of the production process I called *strategy* and *tactics*, which I am here calling *coding* and *grammar*. The interested reader is referred thereto for a more detailed discussion of those stages.

The starting point for the production process is taken to be the decision of some non-linguistic planning process that the most effective way to achieve some goal is by saying something.[1] We imagine then that some specification of this goal, e.g. that some addressee is to be made aware of some state of affairs, together with some model of the mental state of the addressee, is delivered to the linguistic part of the cognitive system. The linguistic system then proceeds, via the stages we are about to discuss, to produce the required utterance(s). Not all utterances originate in the planning process - expletives and such are at least not planned in the same way as more extended forms - but they all presumably funnel into the later stages of the assembly line.

The *framing* stage has a somewhat nebulous but nonetheless crucial task - it must partition the goal it has been given into linguistically manageable chunks, and further characterize these as to propositional content and communicative intent. Exactly how much of the chunking is really a *linguistic* task is unclear, but it seems plausible that at least some linguistic knowledge is involved in e.g. breaking a story up into sentences, so for the time being we will consider the framing task as part of the linguistic system. The framing stage must also further decompose the propositional content into various sub-categories, such as participants, process, modality, and so forth.

The *coding* stage must take the decomposition of each chunk and encode each sub-part thereof with respect to some lexicon, considering as well the model of the mental state of the addressee. In particular the communicative intent must be encoded as a particular speech act, referring expressions for each of the participants must be tailored to the addressee, the process must be encoded, typically as a verbal complex, and so on. Here we see the first possibility of

[1] Cohen [1978], Perrault and Cohen [1977], Allen, James [1979], and, from quite a different perspective, Wilensky [1978], discuss this issue of planning leading to the specification of speech acts.

true parallelism - for the most part the attempt to construct lexical realizations of the participants and the process can proceed independently of one another, and the results be passed to the grammar as they arise. What I mean by true parallelism can be understood in terms of the assembly line image. The task of attaching a wheel to a car can be specified independently of which wheel is involved. With sufficient resources, parallel execution of this task is possible, with all four wheels being attached at the same time, independently. Just so, the lexicalization of a referent as a noun group can be specified independently of which of several referents in an utterance it is, and so with sufficient resources the process of lexicalizing all the referents in an utterance can be carried out at the same time.[2] As each independent lexicalization process finishes, it passes its results to the grammar stage.

The *grammar* stage can thus be seen as receiving over time a stream of groups of lexical items produced by the coding stage, together with other, non-lexical information about the constituents these groups encode, such as the case roles or thematic status of participants, and the tense and aspect of the verbal group, as well as the chosen speech act. Its task is to introduce both strict temporal order and all the necessary grammatical structure into this stream. Articles, particles, agreement markers, case marking, voice marking, mood specification, and so on are all introduced at this level. As soon as each constituent is complete, it is passed on to the phonology stage.[3]

[2] Whether this is always true is not clear. The consequences for this position of a sentence such as *Maureen gave her son his allowance* are complex. On the one hand we can assume a sort of grammatical treatment of at least some pronominalization, in which case the independence hypothesis can stand. In that case, the lexicalization processes yield *Maureen*, *Maureen's son*, and *Maureen's son's allowance*, and it is their order in the utterance which determines the pattern of pronominalization, which would occur in the grammar stage. On the other hand, we could assign pronominalization wholly to the lexicalization stage, in which case the independence hypothesis would have to be abandoned, or at least substantially modified. This choice reflects an ongoing debate on whether there is only one kind of pronominalization, or two - cf. [Reinhart 1976], [Sag 1976].

[3] There is clearly an overly strong commitment to strict decomposition here - in particular we know that the first part of a partially specified constituent may be articulated while the rest is still uncoded, as when someone says *I'll take the ... red one.*

4.2.3 Input and output characteristics of the footmaker

The footmaker, which is *ex hypothesi* the first sub-stage of the phonology stage, thus receives as input an ordered sequence of lexical items. They arrive in groups, with the typical grouping being one constituent, although as noted above smaller groupings may occur, and perhaps larger ones as well. We can think of an input queue, into which groups of lexical items are inserted from the right, and out of which the footmaker removes them as it breaks them down into syllables and assembles them into feet. Exactly what is present in terms of additional structure beyond the simple left-to-right sequence of the lexical items will be discussed below in section 4.3.

What exactly are these 'lexical items' which make up the input to the footmaker? Clearly, they are not strings of characters which give the spelling of the word in question. Literacy is not a prerequisite to speech. Nor does it make much sense to suppose they are copies of the entire entry for a given morph as it appears in the mental lexicon. Such a load on memory capacity is unnecessary. Rather they should be considered as pointers or indices into the lexicon. Such pointers give access to the information in the lexicon as needed by each stage, while avoiding carrying the whole entry along from stage to stage. Previous stages have made use of semantic and/or grammatical aspects of the entries in that lexicon. The footmaker is probably the first stage which makes use of the phonological information which needs must be there as well. In particular sufficient information must be present to enable determination of the abstract stress pattern of the word in question.

Given that the domain of the foot is sequences of syllables, the output of the footmaker must be a sequence of syllables divided into upbeats and feet. Whether output is on a syllable by syllable, foot by foot, or constituent by constituent basis is not clear. At this stage in the development of the theory, little seems to turn on this choice. Many if not all grammatical features of the utterance may be presumed to be discarded (that is to say not passed on) at this point, although features specifying highlighting and tone group related information (boundary location, boundary tones, kinetic tones, envelope, etc.) are of course preserved, as well as any other syntactic or semantic information which is demonstrably needed for the conditioning of subsequent phonological or articulatory processing.

4.2.4 What comes after: Phonology and articulation

I have very little to say about these stages, although the broad outlines are clear - phonemic representation must be translated into motor programs and those programs must be executed to actually produce sound. There are obviously interactions between phonology and the footmaker which I am not considering here - to name two, for example: Morphophonemic variation and fast speech rules both affect syllable structure, and I have ignored issues of reduction and contraction heretofore as well. Whether or not these processes can be successfully specified and integrated into the framework as I have specified it here is moot - my claim is only that the foot-making processes I am concerned with must have a place in any production model. It may be for instance that fast speech rules in some circumstances affect foot structure in ways incompatible with this ordering. But I repeat the qualification made at the beginning of this section: The partitioning I have proposed here is intended as a starting point for discussion. I have attempted to lay out a set of intuitively plausible sub-tasks of the production of an utterance in an intuitively plausible ordering. This idealization of independent stages is a necessary first step from the methodological point of view I have adopted. Once they are formally and completely specified, the task of integration may begin, albeit it may involve fairly drastic restructuring of the independently structured stages, with interactions between stages more complex than the simple, linear input/output model assumed here.

4.3 The footmaker

This section specifies the footmaker in detail. It begins with a discussion of the representation of word stress. Then the footmaker itself is presented in successive stages, mirroring the structure of section 4.1 above.

4.3.1 The Liberman/Prince approach to stress and salience

For a start, we need a specification for word stress which will stand as the basis for determining the division of utterances into feet. By far the most comprehensive proposal along these lines to date is that of Liberman and Prince [1977] (hereafter LP), a development of [Liberman 1975]. I will adopt here their proposals for the analysis of stress at the word level with very little modification, but will not adopt their parallel proposal for analyzing phrasal and clausal stress, which in any case has no place in the approach I am propounding.

4.3.1.1 Metrical trees and the English Stress Rule

To every polysyllabic word LP assign a metrical tree which determines its stress. This tree is binary, with each node establishing a *weak/strong* relation between its descendants. The tree for a given word is determined by the right to left application of a set of principles which first characterize each syllable as ±Stress, and then include it in the metrical tree. These principles are closely related to the English Stress Rule of Chomsky and Halle [1968], with some improvements. The examples below (taken from LP) illustrate the sorts of metrical trees which are assigned.

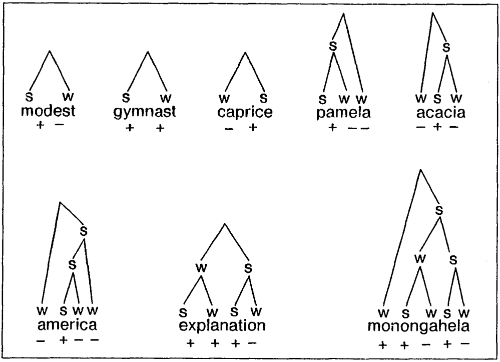
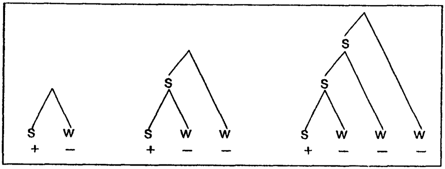
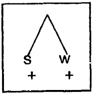


Figure 1. Exemplary metrical trees

In conjunction with the above examples the following summary is offered of how LP uniquely specify a metrical tree for a word. The fundamental well-formedness constraint on metrical trees is that *s* (for strong) may not immediately dominate a syllable which is –Stress. The procedure for constructing a metrical tree is as follows. Proceeding from right to left, first we group together all sequences of the form +–, +––, +–––, etc., and assign them the only tree possible obeying the constraint, namely a left branching trochaic structure:



Sequences of ++ are also joined as trochees:



Finally, stray syllables at either the beginning or the end are joined on, and higher level structure is added. If the fundamental constraint does not determine which way to link the two nodes in this case, it is done so that the right hand of the two nodes is strong if and only if it branches. Compare *execute* with *rehearsal*, and both to *execution*.

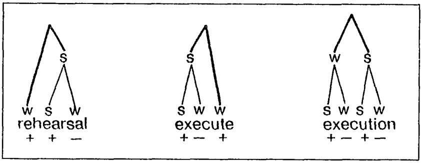


Figure 2. Examples of tree construction

This covers a substantial fraction of English stress. Various additions are necessary to account for various morphologically and historically based exceptions, but a detailed specification of these is not needed here. Suffice it to say that all the word level metrical trees given below are generable by LP, unless otherwise noted.

4.3.1.2 The Relative Prominence Projection Rule and the metrical grid

The metrical tree associated with a given word does not immediately manifest its rhythmic structure. To determine the rhythmic structure, we start from the relational import of the strong/weak pairings: Material dominated by *s* is metrically stronger than material dominated by *w*. From this we can immediately determine which the metrically strongest syllable of a word is. We start at the top of the metrical tree, and follow the *s* branches down until we reach a syllable. This syllable (called by LP the Designated Terminal Element, or DTE, of the tree) is metrically strongest, and corresponds to the primary stress or 1-stress syllable of other approaches. Thus the strongest syllables in Figure 2 of the previous section are respectively *her*, *ex*, and *cu*.

We can extend this approach to determine the relative strength of all the syllables of a word. A syllable has strength proportional to the strength of the material it is stronger than. In a disyllable, the *s* terminal of the *s-w* pair is stronger than the *w* terminal - this is the definition of this relational representation of stress. In polysyllabic words with more structure, the higher levels of structure impose further relations. For instance, in *execution*, as show in Figure 2 above, the second *s-w* pair is in some sense stronger than the first, and this means *cu* is stronger that *ex*. LP state this as their *Relative Prominence Projection Rule* - a requirement that for each node in the metrical tree, the DTE of the strong side sub-tree must be stronger than the DTE of the weak side sub-tree. As stated above, the DTE of a given node is that syllable arrived at by descending from that node by always taking the strong branch. For example consider the metrical tree for the word *reconciliation* (with the non-terminal nodes and syllables labelled for easy reference):

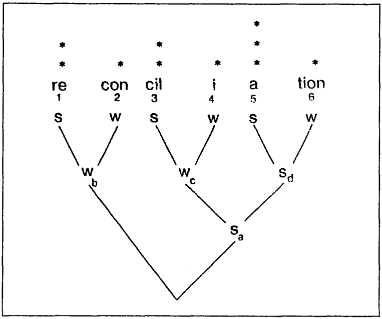


Figure 1. Metrical tree for *reconciliation*

Starting from the head of the tree, the principle propounded above requires first that syllable 5, the DTE of node a, be stronger than syllable 1, the DTE of node b. At node a, the principle tells us that syllable 5, as the DTE of node d, must be stronger than syllable 3, the DTE of node c. And at the syllable level, 1 must be stronger than 2, 3 stronger than 4, and 5 stronger than 6. If the heights of the columns of asterisks above each syllable are taken to indicate the strength of that syllable, then the pattern given above satisfies all the 'stronger than' requirements just listed. It is not the only pattern which would do so, but it is in some sense the *minimal* pattern which would. This pattern may be arrived at for any word in a straightforward way. Start out by assigning one asterisk to all syllables. Then proceed up the metrical tree, at each level assigning additional asterisks as needed to adhere to the RPPR. The result will be a pattern of alternation which LP call the *metrical grid.* The implication is that the grids for individual words are fitted together into one large rhythmic structure, much like the measures of a musical phrase.

4.3.2 The footmaker - simplest version

With this appeal to the parallel between speech and music, we are close to being able to see how to specify the operation of the footmaker in terms of the structures provided by LP. Just as the position of a note within a musical measure gives it intrinsic prominence, either more or less than that of the other notes therein, so the prominence pattern specified by the metrical grid for a word indicates how it must be related to the bar lines, or rhythmic patterning, of speech, which is foot structure.

In most western music, the most common metrical patterns have two, three, or four beats to the bar, with the first beat of the bar the most prominent. As noted in chapter 3, most feet in spoken English have either two syllables (trochees) or three syllables (dactyls), and the first syllable salient. We can think of the task of the footmaker as much like that of the composer with libretto in hand, about to set the words to music. He has a sheet of music paper in front of him, divided into measures but as yet lacking a time signature. He must take the syllables of an utterance and write them under the staves, so that the metrical grid of each word accords with the intrinsic pattern of salience imposed by the bar lines. Just so the footmaker divides the utterance into feet, with the salience pattern implicit in the foot structure according as much as possible with the prominence pattern of the metrical grids. The footmaker has one advantage over the composer: It can assign a time signature to each bar (that is, foot) independently, alternating 3/4 and 2/4 (dactyls and trochees) as necessary, with occasional bars of 4/4 or 1/1 as it were (qsfs and msfs).

To see how this would work, let us consider the following extract from the data: *there was a / lot of expec / tation and ex / citement on / my / part.* If we simply compute the metrical grids for each of the words in this sentence and line them up, this is what we get:

```
                         *
                   *     *            *
  *    *  *  *  *  *  *  *  *   *  *  *   *   *  *   *
there was a lot of expectation and excitement on my part
```

Figure 1. Basic grid for sample utterance

It is clear that some changes have to be made before we can read the foot structure off this grid pattern. The basic problem is that the height of the highest column for each word depends to some extent on the number of syllables in that word. In particular one-syllable words appear to have no more prominence than the weakest syllables of polysyllabic words. What is needed is a normalization, so that the maximum heights are all the same. In addition we need to take account of the long-standing generalization about content versus non-content words. **F1** below specifies how this normalization is to take place. It is preceded by **F0**, which specifies the first step implied by Figure 1 above. Throughout the rest of this chapter we will accumulate and revise these specifications of the steps which constitute the footmaker.

F0 – Associate with each word in the utterance a metrical tree and metrical grid as per LP.

F1 – For all words which are +Salient, normalize their metrical grid as follows:

1) Make the height of the column over the DTE of the word equal to 4. Take the DTE of a one syllable word to be that syllable.

2) Make the height of all columns which are neither over the DTE nor of height 1 equal to 2.

F1 introduces the first of a number of features which we will need to specify the footmaker. The feature ±Salient is predictable on the basis of form class, but does not correspond exactly to the old content/non-content division, nor to the open/closed class distinction, although it is close to the former. There is some correlation with the origin of items of particular form classes within the production system - the lexicon or the grammar. An item whose appearance is totally specified or largely determined by the grammar is likely to be –Salient; one which contributes meaning other than structural meaning is likely to be +Salient. But I have used the name ±Salient rather than, say, ±Gram to emphasize that I have determined actual membership on the basis of foot structure as it occurs, and not on the basis of some abstract criterion. That is to say, each form class was determined to be +Salient or –Salient on the basis of whether its members typically started feet or not in ordinary utterances in the data segment, supplemented by other data when necessary.

The approximate value of the feature emerges from the the following partial specifications of the redundancy rule which specifies which form classes are –Salient, which +Salient:

Article, Conjunction, Copula, (Positive) Auxiliary, Modal, Relative pronoun, Complementizer, Preposition -> –Salient

Noun, Verb, Adjective, Adverb, Quantifier, Negative Auxiliary -> +Salient

Given F0 and F1, we arrive at the following normalized collection of grids for our example sentence:

```
                    *               *               *               *      *
                    *               *               *               *      *
                    *         *     *               *               *      *
   *    *     *     *     *   *  *  *  *     *   *  *   *     *     *      *
there was    a     lot   of expectation    and  excitement   on    my    part
-SAL  -SAL  -SAL  +SAL  -SAL      +SAL    -SAL    +SAL      -SAL  +SAL  +SAL
```

Figure 2. Normalized grids for sample utterance

One of the motivations for the geometric (that is 1 - 2 - 4) rather than arithmetic relation of the column heights established by **F1** is clear here - the underlined portion of Figure 2 has the same pattern as is usually found in a discussion of the rhythmic prominence of the various positions in a measure of 4/4 time ([Liberman 1975], [Potter 1961]). Other reasons will be developed below.

For this sentence we can now read off the foot structure from the grid pattern - if we place a foot boundary in front of every column of height 4 we obtain the desired results. This is formally stated as

> **F2** – Proceeding left to right through the grid, place a foot boundary in front of (before) each syllable with a column of height 4.

4.3.3 Adding the Rhythm Rule

The next phenomenon to be accounted for is that covered by the Rhythm Rule, as discussed above in section 4.1.2. The name 'Rhythm Rule' actually covers two separate problems - the detection of clashing situations, and the relief of that clash by retracting a foot boundary. The first part is called *Clash Detection*; the second *Iambic Reversal.* It is a particular fact about English that a clash will cause a change only if it can be relieved by a shift of a foot boundary to the left - that no shifts to the right occur. Thus we have the clashing *tennes / see / border* shifting to */ tennessee / border*, but we never get */ hard para / dox* from the clashing */ hard / paradox.* Furthermore, the shift will occur even if the result is a clash to the left, e.g. */ west / tennessee / border.* All this suggests that recognition of the clash proceeds from right to left. Convincing evidence for this is provided by the following example, which was recorded at the same session as the data segment: *at the / san francisco / golden gate / bridge*, which in unshifted form would have been *at the san fran / cisco golden / gate / bridge.* The only clash here is *golden / gate / bridge* - it is only after the shift to relieve that clash that a further clash occurs, which then prompts the second shift in *san francisco.*

LP define Clash Detection in terms of the metrical grid and Iambic Reversal in terms of metrical structure. We will consider each of these in turn below. It is important to remember at the outset that these two are separate but interacting processes.

4.3.3.1 Identifying clashing configurations in the metrical grid

LP's definition of a clash is a metrical grid where two asterisks at the same height appear without any asterisk between them at the next lower height. In a normalized grid this means either two 4-columns without an intervening 2-column, or two 2-columns or a 2-column and a 4-column without an intervening 1-column.

In part because of the normalization[4] we will express the rule in terms of column height, rather than asterisk level, and simply identify a clash as occurring under the conditions just listed, which we will refer to as 4-4, 2-2, 2-4, and 4-2 clashes respectively. The status of all but the first of these is somewhat unclear. The 2-2 or 2-4 cases hardly ever occur in any case, as very few words in English end with secondary stress. *Heliotrope* might be such a word, but contrary to prediction, */ heliotrope / seminar* does not seem to clash to me.[5] In the 4-2 case, which is perhaps somewhat more common, there does seem to be some suggestion of a clash, leading to e.g. *tennes / see consti / tution* changing to */ tennessee consti / tution.*

F3A states this approach to clash recognition formally. It computes for each grid position a set of column heights which would clash if they occurred to the left of that position.

[4] This normalization and also the right to left nature of F3 below combine to replace the influence of the supra-lexical level of metrical structure which LP need to condition the operation of the Rhythm Rule

[5] Its tree structure is in any case unclear - perhaps no correction of the clash is possible. As we shall note below, clashes seem to be felt less if there is no hope of relief

F3A – Proceeding from right to left through the metrical grid of the constituent, identify a column as *clashing* on the basis of the membership of the *Clash Set* at that point.

The Clash Set is initially empty (at the right end of the constituent), and its membership is updated at each grid position as follows on the basis of the column at that position:

1) If the column is a 1-column, there is no clash. Remove 2 from the Clash Set and carry on.

2) Otherwise, if the column is a 2-column, then if 2 is a member of the Clash Set, there is a clash. Otherwise, remove 4 from the Clash Set, add 2 to the Clash Set, and carry on.

3) Otherwise, the column is a 4-column. If either 4 or 2 is a member of the Clash Set, there is a clash. Otherwise, add 2 and 4 to the Clash Set and carry on.

This somewhat complex formulation merely attempts to enforce a strict 4-1-2-1-4-1-2-1-4-... organization, while allowing any number of 1-columns to appear contiguously.

Sub-rule 1 of F3A says that when moving leftward, after a 1-column, a 2-column is OK. Sub-rule 2 says that when moving leftward, after a 2-column which itself is OK, a 4-column is OK, provided a 1-column intervenes. And sub-rule 3 says that after a 4-column, we can have neither a 2-column without an intervening 1-column, nor a 4-column without an intervening 2-column. The difficulty of stating this rule concisely when it seems easy to state in visual terms of patterns of columns stems from the necessity of serial exposition. A specification in terms of pattern matching, while closer to our intuitive understanding and taking more advantage of the possibilities of parallelism, would in fact be less efficient.

The right to left nature of this rule seems to be in conflict with the left to right constraint propounded above. But if as hypothesized above in 4.2.3 the footmaker operates on a constituent by constituent basis, and thus divides only small groups of words into feet at any one time, this problem goes away, and in fact something else comes for free. If the footmaker operates left to right on the constituents of an utterance, as they are delivered by the grammar stage, this satisfies the left to right constraint. If it then operates right to left within each constituent, this not only captures the right to left nature of Clash Detection, as discussed above, but it also captures the fact that it rarely if ever operates across constituent boundaries. To accomplish this F0, F1, and F2 above must be modified slightly to apply to ... *each word/syllable in the current constituent ...*

Note also that it is this constituent by constituent style of operation which provides for the variability of foot structure associated with pauses, as described in 4.1.5. As for the lengthening of pre-pausal syllables which figured in Chapter 3, this presumably arises when the footmaker is unable to deliver a foot to the rest of the phonology in time for seamless articulation with the previous one. I suppose the articulation stage in such a circumstance prolongs the last syllable it has for up to an additional tenth of a second in hopes, as it were, of bridging the gap until the next foot.

4.3.3.2 Relief from clashes via Iambic Reversal

The Rhythm Rule is the name given to the situation where there was a clash, and something was done to relieve it. Notionally what happens when a clash is detected is that a syllable with secondary stress early in the word is promoted to primary status, while the erstwhile primary-stressed syllable is demoted to unstressed status. There are some subtleties to this process, and LP have demonstrated that it is best to formulate the process in terms of the metrical tree of the word involved. They call it *Iambic Reversal*, as what is involved is taking a *w-s* pair in the metrical tree and reversing it to *s-w*. If the *s* of the original pair dominated the clashing syllable, this reversal has the effect of moving the DTE of the word away from that syllable, and thus relieving the clash by requiring a re-computation of the metrical grid associated with the word. Before discussing restrictions on this reversal, or specifying it formally, I will exemplify both the operation of F3A above, and the effect of Iambic Reversal, with respect to the constituent *an / anaphoric / reference.* Figure 1 below shows the grid for this constituent, and the successive states of the Clash Set from right to left, up to the point where the clash occurs.

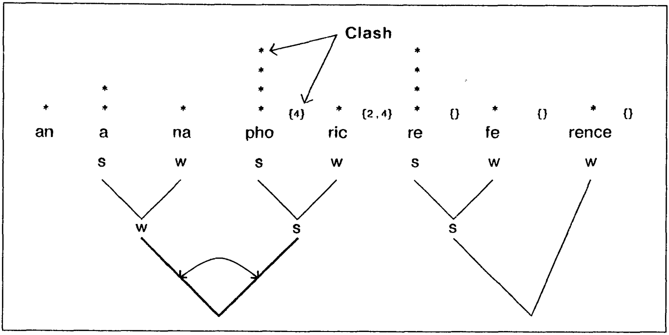


Figure 1. Clash computation for *an anaphoric reference*

If the indicated *w-s* pair is reversed and the grid recomputed, the result is as in Figure 2, with no clash.

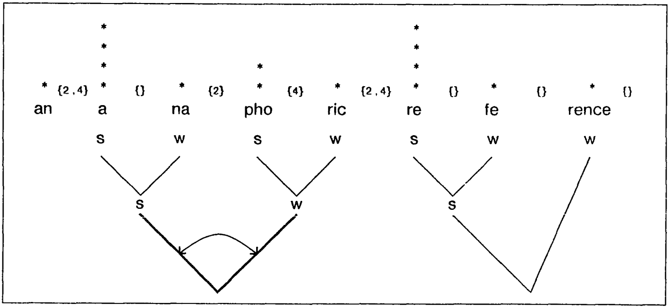


Figure 2. Result of Iambic Reversal in *an anaphoric reference*

Iambic Reversal may not apply to any *w-s* pair without restriction. There are two constraints on its operation. First of all the fundamental well-formedness principle for metrical trees applies.

The metrical tree resulting from the reversal may not have the retracted *s* node directly dominating a syllable which is –STRESS. For example compare *tattoo* with *canoe.* Both have a *w-s* metrical tree, but the first syllable of *tattoo* is +STRESS, while that of *canoe* is –STRESS. The consequence of this is that Iambic Reversal may apply to *tattoo,* as in *the / tattoo's / color,* but it may not apply to *canoe,* so we do not get e.g. ** the / canoe's / owner.*

The other constraint has to do with the location of information focus within the constituent, a concept discussed briefly in 4.1.6 above. LP state it as blocking a reversal which would change the "DTE of an intonational phrase [our *tone group*]" but go on to suggest a broader restriction. As our system does not have metrical structure above the word level, this formulation is not directly translatable. I will state the restriction in terms of information focus, which is a concept of both broader scope, in that it applies at different levels, not just that of the tone group, and greater content, in that it is based on semantic and pragmatic considerations rather than formal structure. The detailed nature of these differences will become clear as the use to which I intend to put this notion becomes clear.

Before discussing it however some problems with this approach to the Rhythm Rule need to be considered.

Two aspects of the proposed formulation of the Rhythm Rule in terms of Clash Detection and Iambic Reversal as specified above might lead to some doubts. First, the multiple application of the Rhythm Rule can be carried to extremes, as in */ thirteen / good-looking / antique / tennessee / raccoon / coats.* I confess to a certain unease at the thought of right to left application over such a large domain. Second, there are the times when the rule does in fact apply across a constituent boundary.

One approach to these problems is to point out that Iambic Reversal can apply without a clash having been detected. In some dialects or speech styles, this may in fact be the norm. Some people may simply always say */ tennessee* in all but utterance final or citation contexts, and in fact I've even heard it that way utterance finally in a Country and Western song. Overzealous application of Iambic Reversal is not a problem then - what *would* be a problem would be consistent failure to get Iambic Reversal in cases where there *was* a clash, but neither of these so-called problems is like that.

On the other hand, for people like me who by and large seem to do Iambic Reversal only when there is a clash, we still need an explanation for these two problems. In the cases when Iambic Reversal occurs as a result of a clash across a constituent boundary, we can argue that some unit larger than a constituent was delivered to the footmaker, and it does in fact seem that most counter-examples depend on over-learning, which would presumably enable a whole utterance to be processed as a unit. For example I claim that if you chance to utter the sentence *Tennessee borders Kentucky*, and you share my dialect, you will say it as *Tennes / see / borders Ken / tucky*, but after a few repetitions, the Rhythm Rule may intrude to yield */ Tennessee / borders Ken / tucky*, as the sentence comes to be produced as a whole, and the constituent boundary 'goes away' as far as the footmaker is concerned.[6]

The issues of reading and over-learning mitigate the multiple application problem somewhat as well. In general, there is a problem in investigating the Rhythm Rule via constructed examples such as this one. People's conscious predictions about when and where they will or will not apply it are often wrong, and in this regard reading aloud is particularly problematic. Reading aloud and also the over-learning which accompanies repeating back a sentence several times obviously change the situation with respect to the Rhythm Rule, in particular with respect to the size of the unit of production. The sort of fore-knowledge which a speaker has in these cases is quite different from that available in fluent, unrehearsed conversation. And, unfortunately, the rate of occurrence of the relevant phenomenon is extremely low in ordinary conversation, that is, two or more words in sequence, within a single constituent, pronounced without a break, all liable to Iambic Reversal. The *San Francisco Golden Gate Bridge* example above is the only time in about three hours of taped conversation which I have scanned where such a situation arose. In any case the current proposal elegantly covers the common cases, and much more live data will be required to diagnose the refinements needed to accurately cope with the pathological cases.

[6] Such a constituent-final application of the rule as this also violates the +FOCUS constraint to be discussed below, but as we assume this feature to be redundantly supplied on constituent-final items, the reanalysis of the utterance as one large constituent leaves Tennessee with no +FOCUS feature, so there is no problem

4.3.4 Infonnation focus

As noted above in 4.1.6, the information focus in a unit, whether noun group, verb group, or clause[7] is that sub-constituent bearing the principal communicative content of the unit. Typically this is the last or rightmost sub-constituent. This idea is well established at the clause level - I am proposing extending it to smaller units as well. The phenomenon in these smaller groups is not identical to that at the clause level, but it is close enough I think to justify a unified analysis. Thus the last noun group of a clause is typically the bearer of new information, the head noun of a noun group typically is the principal determinant of the referent of the group, and the finite verb is the principal specifier of the process encoded by a verb group. If we take the function of the clause to be informing, of the noun group to be referring, and of the verb group to be specifying process, then for each of these in the unmarked case it is the rightmost sub-constituent which bears the principal functional load, which makes the greatest contribution to the function of its parent.

Kinetic tone placement is sensitive to focus, so in the typical case we have the kinetic tone falling on the last group of the clause, and on the last sub-constituent therein, and so on.

But rightmost focus is just the unmarked case. For instance, the verb group or even the subject noun group may be the new information, and thus focused, at the clause level, and a prenominal modifier may be the principal determinant of reference, and thus focused, at the noun group level. This is in fact usually the case with Noun-Noun compounds and measure phrases such as *eight hour,* and sometimes happens with Adjective-Noun structures as well. Thus we distinguish two cases - the unmarked case, in which the last element is the information focus, which we will mark as +FOCUS, and a marked case, in which some non-final element is +FOCUS, and the subsequent elements –FOCUS. Formally, we add a redundancy rule as follows:

[7] In fact focus is specified in the domain of information structure, not syntactic structure, so this is something of a category error, However, the units often coincide, so for the time being I will speak of focusing within groups and clauses. Section 4.3.7.2 takes up this point in more detail.

F4 – Within any group level domain,[8] if any non-final immediate constituent of that domain is marked +FOCUS, mark all immediate constituents to the right of that one -FOCUS, otherwise, mark the final immediate constituent +FOCUS.

There are three things we want this marking of ±FOCUS to accomplish for us. These are the interaction with Iambic Reversal noted by LP, the effect on foot boundaries exemplified above in (2) of 4.1.4, and the effect on kinetic tone placement. Here we consider the interaction with Iambic Reversal - the other effects will be addressed in subsequent sections.

4.3.4.1 The interaction of focus and Iambic Reversal

LP observe that Iambic Reversal sometimes fails to apply in rhythmically clashing Noun-Noun compound situations, so that we don't get (1a) shifting to (1b):

1a) *an / tique (/)*[9] *dealer*
1b) */ antique / dealer,*

at least not with the same meaning. That is to say, (1a) refers to a dealer in antiques, where (1b), if it is possible at all, refers to a dealer who *is* antique. LP attribute this to the fact that in the Noun-Noun reading of (1a) *antique* is the DTE (in our terms +FOCUS) of the intonational phrase, and thus Iambic Reversal is blocked. Thus (1b) can only have an Adjective-Noun reading, where *dealer* is the DTE. Then they contrast

2a) *kanga / roo (/) rider*
2b) ** / kangaroo / rider*
2c) */ kangaroo / rider's / saddle,*

with (2a) and (2b) parallel to (1a) and (1b), and claim that where *kangaroo* is no longer the DTE for the whole noun group, as in (2c), it is free to undergo Iambic Reversal. But on my account it still may be +FOCUS within its subordinate group, and the operation of Iambic Reversal is possible only if it is not, whereas the LP formulation implies that no such semantic difference is necessarily implied. I think this is an oversimplified view, which stems from trying to account for things in purely structural terms. I think their acceptance of (2c) with a Noun-Noun reading coupled with their rejection of (2b) can be viewed in two ways - the conservative

[8] We leave out the clause level = the tone group as that needs must be handled separately, as the footmaker typically never sees a whole tone group at once.

[9] This and subsequent parenthesized foot boundaries are optional. See section 4.3.4.2 below for some discussion.

and the liberal, as per the discussion above in 4.3.3.2. In the conservative dialect Iambic Reversal is strictly conditioned by clashes and constrained by +FOCUS. The contrast between (2b) and (2c) is explained by saying that in fact in (2c) within the sub-group *kangaroo rider's* the word *rider's* is in fact +FOCUS, but that since there is no sensible Adjective-Noun reading we get the Noun-Noun reading anyway - which is only to say that semantic considerations can sometimes override phonological cues, which in any case are only heuristic at best. The consequences of this conservative position become more clear if we considering another of LP's examples, where they only gave (3a) and (3b), parallel to (1a) and (1b), without commenting on the interpretation of (3c) and (3d):

3a) *Chi / nese (/) expert*
3b) */ Chinese / expert*
3c) */ Chinese / expert's / salary*
3d) *Chi / nese (/) expert's / salary*

I claim that (3c) strongly favors the reading of (3b), referring to the salary of an expert who is Chinese, while (3d) parallels (3a) in referring to a expert whose speciality is Chinese. To the extent to which this is the case, the conservative viewpoint is upheld, and the assignment of +FOCUS at any level can prevent the operation of Iambic Reversal.

In the liberal dialect, on the other hand, Iambic Reversal operates more or less independently of any conditioning or constraint. For such a dialect there is not really any absolute prohibition on Iambic Reversal applying in the presence of +FOCUS. Many people feel, for instance, that (2b) above is just fine, and Lisa Selkirk (personal communication) reports a number of examples with the kinetic tone located on a syllable which had been promoted by Iambic Reversal. Since kinetic tone follows +FOCUS, this is clear evidence of Iambic Reversal occurring despite +FOCUS. We are thus led to the conclusion, at least for some dialects, that the presence of +FOCUS has little or no inhibiting effect on Iambic Reversal.

In any case, if there *is* a second condition on Iambic Reversal, then I want to state it in terms of +FOCUS, where that feature is articulated at any level of structure, for example, as above, with respect to possessive noun groups.

All this leads us to the following formal specification of Iambic Reversal and its interaction with clash detection as given in **F3A** above:

F3B – A *w-s* pair whose DTE is the current syllable may be reversed to *s-w* provided that

1) the new *s* does not immediately dominate a syllable marked –STRESS.

2) (the current syllable is not marked +FOCUS.)

If the current column clashes, there is a strong pressure in favor of performing this reversal if possible.

If the reversal does take place, the latter half of FO (grid calculation) and F1 are reapplied to the word containing the current syllable, and F3A starts over at the right hand end of the word.

where the second condition may or may not be relevant, as discussed above.

Note that this formulation does not tie the operation of Iambic Reversal strictly to the detection of a clash - it may happen without one being present, as in / *I like* / *antiques*, which is odd but not impossible. On the other hand, it need not happen when there *is* a clash, as in 4.1.3(1c) above, although it is strongly favored.

Also note that no specification is made in F3B of *which w-s* pair is to be reversed, in cases where more than one might be eligible, e.g. in *onomatopoetic*, which has the following metrical tree:

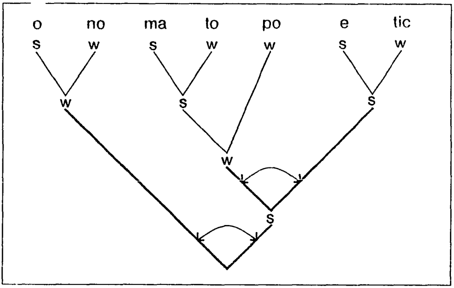


Figure 1. Metrical structure for *onomatopoetic*

In fact on asking people to read the sentence *He showed a tendency for onomatopoetic* *passages* I got */ onomatopoetic / passages* from some, and *ono / matopoetic / passages* from others (as well as */ onomato / poetic / passages*), suggesting that at least in some cases either *w-s* pair may switch, although which is chosen might in fact on further investigation prove to be consistent for a particular individual.

4.3.4.2 The effects of –FOCUS*: Bleaching*

Two of the circumstances in which primary stressed syllables of +SALIENT words are not realized as salient mentioned so far have a lot in common. These are bleaching following highlighting, discussed briefly in 4.1.4 above, and the effect on foot structure of Noun-Noun compounding, mentioned in 4.1.6 above, and again in the preceding section. I propose to unify these two phenomena, and attribute them both to the presence of –FOCUS. The unification part is easy, in that given the meaning of information focus, it is clear that most if not all highlighted constituents are information foci, and thus +FOCUS, while subsequent sub-constituents within their group are –FOCUS. It seems that if we could arrange for **F2** to be sensitive to –FOCUS, and not start a foot where it otherwise would, we would be on the right track. But the situation is not quite so simple. Consider the following examples:

1a) *I / said it / ↑depressed me / not that it / ↑impressed me*
1b) */ steel warehouse*

2a) *I said / ↑a con (/) dition not / ↑the con (/) dition*
2b) */ copper (/) warehouse*

3a) *↑a conso / lation not / ↑the conso / lation*
3b) */ copper po / lice*

The bleaching effect seems to wear off - quite quickly in fact. Only the immediately following syllable seems to be affected for sure. The position of word boundaries also seems to have a slight effect - I think I'm more likely to say (4a) than (4b), but (5b) sounds better than (5a):

4a) */ copper / panel*
4b) *? / copper panel*

5a) *? / steel com / mission*
5b) */ steel commission*

Also in this regard note the example of *↑I remember just / being ...* from the text.

Some suggestion of how to proceed comes from the comments in section 4.1.3 about the pressure for alternation. If we think of there being a threshold for starting a new foot, which is high immediately after starting a foot, and which declines after each syllable past the start, dropping to a low level after three or four syllables, we not only have a rough characterization of the pressure for alternation, we also have a refinement of F2 which will enable us to deal with the bleaching associated with -FOCUS, and other things as well.

We set the threshold at height 4 immediately after starting a foot, and decrease it by 1 after each subsequent syllable that does not itself start a foot. A new foot is started whenever the column height of a syllable is greater than or equal to the threshold.[10] If we now let -FOCUS have the effect of reducing column height by one, the desired effects follow. Consider the following diagrams of (1b), (2b), and (3b), where the threshold is given below each column:

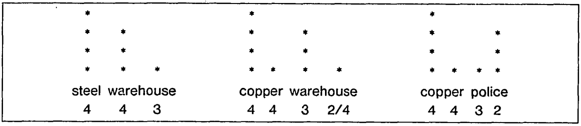


Figure 1. Columns and thresholds for some exemplary compounds

In each of these examples, the first syllable starts a foot regardless of the threshold, as it is a 4-column. This sets the threshold at 4 for the next syllable. The three examples differ in where the 3-column, reduced from a 4-column by the effect of -FOCUS, is located with respect to the declining threshold. In *steel warehouse*, it is in first place. 3 is less than 4, so the -FOCUS primary stressed syllable *ware* immediately following the salient syllable *steel* will probably not start a new foot. In *copper warehouse*, the 3-column is one syllable further on. 3 is equal to 3, and we may or may not get a new foot. And finally in *copper police*, the 3-column is two syllables away. 3 is greater than 2, and we probably will get a new foot.

This proposed reformulation of F2, which is sensitive to differing column heights, has other good effects. Disyllabic prepositions have a somewhat anomalous status. They are -SALIENT,

[10] We can use the distinction between the *greater than* and *equal to* cases to fine tune, saying that the chances are very good for a new foot in the *greater than* case, pretty good in the *equal to* case, and bad otherwise. Also note that this variation is not possible at the endpoints: A 4-column always starts a foot and a 1-column never starts a foot, regardless of the threshold level.

but sometimes begin feet, as noted in 4.1.3, either at the beginning of tone groups, or after a number of unstressed syllables. This now follows, provided we start with a threshold of 2 at the beginning of tone groups, since **F0** will give a 2-column/1-column structure to e.g. *under*, and **F1** will leave that alone, as prepositions are –SALIENT.

One further refinement seems called for in keeping with the relational nature of things. We shouldn't start a foot with less than a 4-column unless there is a subsequent column, and it is lower. If it's equal or higher, we'll get a better rhythmic pattern by waiting and starting the new foot with *it*. This accounts for the preference for (6a) over (6b), where *mine*, having only a 3-column as the result of being –FOCUS, tends not to begin a foot as there is no lower column following it.

6a) */ copper mine*
6b) */ copper / mine*

All this gives the following addition to **F1** and reformulation of **F2**:

F1 - addendum

3) For each syllable in the word which is marked –FOCUS, decrease the height of its column by 1, unless it is already a 1-column.

Note that this implies that all syllables of a word which is –FOCUS are themselves –FOCUS. Feature propagation is discussed below in 4.3.7.

F2′ - Set the threshold at 2. Proceeding left to right through the constituent grid, at each syllable

1) if the associated column is a 4-column, start a foot with this syllable, set the threshold at 4, and carry on

2) otherwise if the height of the column is greater than or equal to the threshold, and there is a next column whose height is less than the height of this one, either start a foot with this syllable, set the threshold at 4 and carry on, or else decrease the threshold by 1 and carry on. Starting a foot is preferred if the column height is greater than the threshold.

3) otherwise decrease the threshold by one and carry on

Note that in no case will a 1-column start a foot, as the condition requiring the next column to be lower in height can never be satisfied. Also note that by parameterizing the amount by which the threshold is decreased at each syllable, we can distinguish various levels of speech rate to the extent it can be controlled at the purely linguistic level.[11] The system as specified, with a decrease of 1, is designed to give 'normal' behavior. If we decrease the threshold faster, say with a drop of 2 for each syllable, this will allow feet before virtually all secondary stresses, yielding an exaggeratedly slow, rhythmic, and careful style. If on the other hand we decrease the threshold more slowly, with a fractional drop of less than 1 for each syllable, this gives a high speed style, with relatively fewer feet, each consisting of more syllables.

4.3.5 More on prepositions and the pressure for alternation: The Rhythm Rule revisited

The previous section dealt with the ambiguous status of disyllabic prepositions and offered an explanation for their occurring sometimes as salient, sometimes not, in terms of the possibility of starting a foot with less than a 4-column, as per **F2'**. It is less clear how to deal with monosyllabic prepositions and verbal particles. If we compare (1a) and (1b) below with (2) in 4.1.3 above (repeated here) we see that the situation is not quite the same:

1a) *The / books on the / table are / good.*
1b) *The / newspapers / on the / table are / good.*
1c) *The / newspapers on the / table are / good.*

2a) *The / books under the / table are / good.*
2b) *The / newspapers / under the / table are / good.*
2c) *? The / newspapers under the / table are / good.*

I am pretty sure that there is more than just a phonological difference between (1b) and (1c), that more semantic significance accrues to *on* in (1b) than in (1c). Also, compare (4a) and (4b) below with (3) from 4.1.3:

3a) */ Under the / table / crouched a / bear.*
3b) *? Under the / table / crouched a / bear.*

4a) */ In the / corner / stood a / figure.*
4b) *In the / corner / stood a / figure.*

(4b) is clearly preferred, with (4a) being semantically marked, while (3b) is difficult if not impossible. As noted above in 4.3.4.2, the 2-column associated with the first syllable of *under* is sufficient to start a foot utterance initially, while the 1-column of *in* is not. Again, as in (1b) and

[11] The principal determinant of speech rate, the ratio of pause to speech, is primarily non-linguistically controlled, in that it is determined mostly by the framing stage and other cognitive stages prior thereto.

(1c), there is a sense that the *in* in (4a) conveys significant content, while in (4b) it is merely grammatical. It is important to understand when considering these examples that no highlighting is going on here - there is no particular tonal prominence associated with *in* in (4a) beyond that small amount which normally accompanies a foot boundary.

All this suggests that there is more to the realization of monosyllabic prepositions as salient than just phonologically based pressures. We will return to this in the next section. But there is another phenomenon involving prepositions which *does* seem to be phonologically conditioned, namely that exemplified in (1) of 4.1.3, repeated here:

5) *at / least all / through that / day*

This is an example of a small class of adverb-preposition compounds which exhibit curious behavior. More examples follow: .

6a) *There were / riots all / over De / troit.*
6b) *? There were / riots / all over De / troit.*

7a) *? There were / riots all / over / Newark.*
7b) *There were / riots / all over / Newark.*

The ?'s in the above may be a bit strong, but there does seem to be a clear preference for (6a) and (7b). The ad-prepositional adverbs which participate in this construction include *all, just, right,* and *straight.* Monosyllabic prepositions exhibit similar behavior when preceded by these ad-prepositions, as in (5) and the following:

8a) *The / climbers / climbed straight / up the / cliff.*
8b) *The / climbers / climbed / straight up / Everest.*

9a) */ Robin / kept right / on with the / work.*
9b) */ Robin / kept / right on / working.*

What's going on here seems to be a restricted version of the Rhythm Rule. The ad-preposition and the preposition are so tightly bound together that they form a single metrical unit with a single metrical tree, and are thus liable to the Rhythm Rule. This means these compounds must have structures like the following:

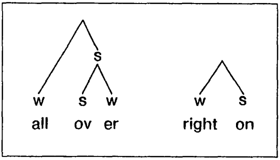


Figure 1. Metrical structures of adverb-preposition compounds

The force of the Rhythm Rule seems somewhat less here than in the simpler cases. In (6) and (7) for instance, the extra 1-column which differentiates (6a) from (7a) should not make a difference to the perception of a clash, according to F3A, but apparently it does. It seems that this adverb-preposition compound structure requires a "stronger" clash before it will retract. The end-stressed prepositions like *before*, lacking the buffering 1-column, behave more according to expectations:

10a) *the / exit / just before / Boston*
10b) *? the / exit just be / fore De / troit*
10c) *the / exit just be / fore Chatt / nooga*

4.3.6 Semantically based promotion and demotion

As noted above in 4.1.6 and in the last section, a word which bears more or less semantic content than usual may be promoted or demoted accordingly in terms of its salience. As was pointed out when it was introduced in 4.3.2, ±SALIENT itself is close to an encoding of the inherent presence or absence of content in words of particular form classes. A consideration of what classes are +SALIENT, such as nouns, verbs, and adjectives, as opposed to those which are –SALIENT, such as articles, copulas, and complementizers, gives a good idea of what is meant by 'semantic content'. Words belonging to classes of the first type usually have it when they appear in utterances, words belonging to the second usually don't. Sometimes the context of utterance leads to a reversal of this expectation, with –SALIENT items being 'promoted' as it were to the status of having content, or +SALIENT items being 'demoted' to non-content status. In the promotion category we have prepositions (with the effect more visible for monosyllabic ones) and pronouns (particularly subject pronouns). All the examples in the previous section described as involving more than just a phonological difference between different foot structures (4.3.5 (1b), (2c), (4a)) are examples of this sort of promotion.

Although I am claiming this is not the same as highlighting, that it is possible to in effect simply to change these words from -Salient to +Salient, without any tonal prominence, nonetheless the two phenomena are probably all of a piece. As noted above in 2.3 the problems subjects experienced with the tonal excursion task probably stemmed from the non-discrete nature of the tonal and rhythmic realization of increased semantic content. Some increase over the normal on these items is sufficient to provoke a foot boundary, and some further is enough for tonal accent. Just how much excursion is necessary to merit the distinction seems to have been the problem. It is relevant in this connection to point out that over 50% of the disagreements in foot boundary marking between subjects H, A, and C were at points where promotion or demotion was involved. This points up the idealization implicit in this whole enterprise - namely that feet are either there or not - there is no middle ground. But in fact we see that there is not always a clear-cut distinction, and the introduction of words like *preferred* in F2′ and *may* in F3B is a reflection of this uncertainity.

In any event, we introduce the feature ±Content as a way of capturing these promotions and demotions. It essentially provides a way of reversing the ±Salient marking of a word. +Content sets the highest column of a word so marked to height 4, and -Content, like -Focus, subtracts 1. -Content will thus have its greatest effect on monosyllabic items at the end of constituents, preventing them from starting feet. This is in keeping with the observations made above that monosyllabic elements are the most likely to drop out. In the text we have *took, take, mean, things, new,* and *know* (twice). Monosyllabic adjectives also frequently drop out when they are redundant or formulaic, as when *old* reduces to *ol'* in such contexts as *i / guess it's / time to / hang up the ol' / six gun and / put ol' / Paint out to / pasture.*

Formally all this leads to another addition to F1:

F1 - addendum

4) If the word is +Content, set the height of the column over the DTE to 4.

5) For each syllable in the word which is -Content, decrease the height of its column by 1, unless it's already a 1-column.

As for verbs with associated particles, I am uncertain as to the correct analysis. When separated, they behave pretty much like ordinary verbs on the one hand, and intransitive prepositions on the other. There is some suggestion that when the particle is not moved, it is sufficiently tightly bound to the verb to be a metrical unit, with one metrical tree, as in */ I would wake / up* from the text, but my data is too sparse and my intuitions too uncertain to be sure this is the right approach.

4.3.7 Feature propagation and kinetic tones

Up until now I have not exactly stated the behavior of the features I have introduced, using them indiscriminately at both the word and syllable level. This section more carefully specifies the nature of each of the four features used by the footmaker.

4.3.7.1 ±SALIENT

±SALIENT is a feature of words, more particularly of lexical items, and is redundantly specified by their grammatical category. A complete list follows, based largely on the patterns of foot boundary assignment in the data segment:

+ SALIENT:	Adjective, Adverb, Demonstrative Pronoun, Intransitive Preposition (e.g. *the / night be / fore*) Noun, Negative Auxiliary (e.g. *that / weren't / opened*), Negative (e.g. *not, never*) , Number, Participle (Past, Passive, Present), Quantifier, Verb, Verbal Particle.
–SALIENT:	Article, Auxiliary, Complementizer, Conjunction,[12] Copula,[13] Determiner, Dummy (*there, it*), Particle (*of* in e.g. */ all of / those, to* in e.g. */ went to / sleep, have to / clean*), Possessive Pronoun, Preposition, Pronoun, Relative Pronoun.

As noted above, however, there is a close parallel between ±SALIENT and some sort of intuitively plausible content/non-content distinction, as a consideration of the above lists will suggest.

[12] Simple one syllable conjunctions such as *and* or *but* are clearly –SALIENT. I am not sure about others like *since* or *because* etc.; more data is needed before a sure determination can be made.

[13] This category includes more than just *be* and *become*. Main verbs functioning as copula also are –SALIENT, as Halliday has pointed out. Compare for instance */ Robin / grew po / tatoes* with */ Robin grew / weary.*

4.3.7.2 ±FOCUS

±FOCUS is a feature not in the propositional/semantic system, but rather in the textual/information structure system. At each level of information structure, it marks the bearer of the principal functional load - informing, referring, etc. As such its proper domain of articulation is over units of information structure, in particular the information unit and its realization the tone group. Usually the information unit is co-extensive with a clause, and the informational sub-constituents of the information unit are co-extensive with the syntactic sub-constituents of that clause, that is with noun groups and verb groups. But not necessarily. Exactly what other possibilities for sub-constituency of the information unit exist is not clear. Selkirk [1979, forthcoming] has proposed the existence of the phonological correlate of these sub-constituents as a necessary element of phonological representation, but largely on the basis of evidence from languages other than English. The only evidence I can offer within English is indirect and in some sense circular - namely the distribution of ±FOCUS itself. It seems natural to suppose, for instance, that the division of the clause into several blocks of given and new material material which accompanies highlighting is reflected in a partition of the information unit into two sub-constituents. The following excerpt from the text can be analyzed in this way:

1) / $\overset{\uparrow}{I}$ *remember just / being just a / super / happy / kind of / time* |

where the first informational sub-constituent covers *I* and *remember*, with marked focus on *I*, and the complement clause is covered by the other informational sub-constituent, with unmarked focus on the long final noun group. Clearly this is an area where more work is needed.

For the sake of discussion, we will consider ±FOCUS in terms of clause and group level constituency. That is, within each clause, noun group, and verb group one element is marked +FOCUS. As specified in **F4** above, in the unmarked case this is the last element in the group or clause, otherwise those elements past the +FOCUS element are -FOCUS. Elements marked +FOCUS may be either complex (rank shifted clauses, hypotacticly related groups/clauses) or simple (words). The feature does not propagate if it is assigned to a complex element, as ±FOCUS will be independently articulated within the domain of that complex element. The feature is propagated from words to syllables as follows: If a word is +FOCUS, then its DTE is +FOCUS. If a word is -FOCUS, all its syllables are -FOCUS. Highlighting implies focus, and since highlighting may apply at the syllable level,[14] we may have ±FOCUS assigned directly at the syllable level, as in 4.3.4 (1a) above. In this case, focusing overrides the intrinsic metrical structure of the word, and the metrical structure is rearranged so that the focused syllable is the DTE.

4.3.7.3 ±CONTENT

±CONTENT is assigned to words or groups. In the latter case it propagates to all words within the group. Propagation from words to syllables is as for ±FOCUS: If a word is +CONTENT, its DTE is +CONTENT. If a word is -CONTENT, all its syllables are -CONTENT. ±CONTENT can be viewed as a way of reversing the unmarked ±SALIENT marking of a word in response to its bearing marked semantic load, either more or less than usual. As such +CONTENT is a sort of early stop on the way to highlighting, and -CONTENT is an early stop on the way to ellipsis. Given the definition of +FOCUS, we would expect to see +CONTENT on any item in marked focus which is -SALIENT.

4.3.7.4 Highlighting

I assume a feature ±HIGHLIGHT to mark the locus(es) of highlighting. It may be assigned at group, word, or syllable (morpheme) level. It propagates from group to word level by following +FOCUS, that is, if a group is +HIGHLIGHT, then its +FOCUS element is +HIGHLIGHT, and if that element is complex, then *its* +FOCUS element is +HIGHLIGHT, and so on to the word level. Propagation from word to syllable level is via metrical structure: If a word is +HIGHLIGHT, then its DTE is +HIGHLIGHT. At the level at which +HIGHLIGHT is actually specified, we expect to see the highlighted element also marked +FOCUS, both because it seems indicated on semantic grounds, given the function of highlighting, and because it has the desired effect of marking the balance of the relevant constituent -FOCUS, which in turn bleaches out closely following syllables, as per 4.3.4.2. We would also expect +CONTENT marking if the highlighted word is not +SALIENT, again both for semantic reasons, and to insure that we get a foot boundary.

The fact that more than one word in what appears to be a single tone group can be highlighted confuses the issue of the relationship between +FOCUS and +HIGHLIGHT, and also

[14] Actually almost always it is morphemic, not syllabic, contrasts which are involved.

provides more evidence for the complicated state of affairs with respect to interactions between the Rhythm Rule and +Focus, as the following example points out:

1a) */ How do you / get to / work?*
1b) *I / take the / San Mateo / Bridge.*

2a) *I / understand you / take the / Dumbarton / Bridge to / work.*
2b) */ No | I / take the San Mat / ↑eo Bridge.*

3a) *I / understand you / take the / Berkeley / Ferry to / work.*
3b) */ No | I / take the / ↑San Mateo / ↑Bridge.*

(1) and (2) show the standard, conservative, pattern. Iambic reversal has applied in (1b), but highlighting and the accompanying +Focus have blocked its application and bleached the foot boundary before *Bridge* in (2b). But in (3b), with two items highlighted, it seems we are back to normal in some sense, with the foot structure being parallel to that in (1b).

It is not clear how to formally account for what is going on here, although it seems intuitively clear that what has happened is that the ante has been raised across the board, and the Rhythm Rule is applying at the resulting higher level. The application of Iambic Reversal precludes division into two tone groups or even two informational sub-constituents. Assigning +Focus to both *San Mateo* and *Bridge* seems inconsistent with the relational nature of ±Focus as I have defined it, and would have problematic consequences for the assignment of kinetic tone, as outlined in the next section. My best guess is that this kind of double tonal accent is an alternative, more emphatic, realization of the highlighting of a complex constituent. As has been noted before, by Chomsky [1971] and others, the sort of proposal made above, namely that +Highlight propagates by following +Focus down to the word level, leads to an ambiguity of analysis in examples where tonal accent occurs on the word carrying unmarked focus in a complex constituent, as in:

4) */ Purple / ↑cows are / funny.*

In vacuo (4) could either be analysed as highlighting *cow* within the domain of the noun group *purple cow*, or as highlighting that noun group within the domain of the whole sentence, with +Highlight propagating to *cow* as the unmarked focus within that noun group. (3b) then can be seen as a way of unambiguously indicating that is is the latter interpretation, with the whole noun group highlighted, which is desired. Presumably the details of this are that +Highlight can propagate either to +Focus or else to all +Salient and +Content subconstituents in extreme cases. This solution leaves e.g. *San Mateo* not marked +Focus, so Iambic Reversal may apply even in the conservative dialect, which seems to accord with such judgements as I have gotten on these and similar sentences. On the other hand, *Bridge* is +Focus as the unmarked focus, and indeed it does seem to bear the kinetic tone, which brings us to the next section.

4.3.7.5 The location of kinetic tone

Similarly to highlighting, kinetic tone follows +Focus down to the word level, and then lands on the DTE. This process starts at the tone group or clause level, but the choice of which intonational sub-constituent or group will be focused at that level is outside the scope of view of the footmaker, since it is restricted to operating on one sub-constituent or group at a time. As mentioned above in 4.2.3 I assume that the input to the footmaker includes both end of tone group indications, and specification of the kinetic tone and the sub-constituent or group it goes with. In the majority of cases, where the tone group is co-extensive with a clause and the groups of the clause are the intonational sub-constituents, the unmarked case will be for the kinetic tone to be associated with the last group with content or new information, which is normally the last group in the clause. This accords with our sense of inconsistency between +Focus and –Content or indeed –Salient. This is the statement in terms of the structure of the footmaker of the oft made claim that the unmarked location for the kinetic tone is the last content or closed class item in the utterance. In this respect it is worth noting that utterances ending with pronouns (–Salient) or anaphoric NP's (–Content) typically do not have focus there, but rather must have it on a previous element, typically the verb. This in turn gives us the kinetic tone on the verb, as in the following examples:

5a) *How was your visit to the dentist?*

            \
5b) *I'd / like to / murder him.*

            \
5c) *I'd / like to / murder the / bastard.*

This is what has been called "anaphoric destressing" ([Ladd 1978]), but in our terms this is not a case of destressing, but if you will of defocusing, with a marked focus shifting the kinetic tone back. The marked focus arises because of the –Content or –Salient marking of the final group. Note in particular that *bastard* does get a foot, although a monosyllabic noun might not.

Another way of looking at this is that there are two ways marked focus can arise - by active choice to mark a particular constituent as focused, or else in reaction to a –CONTENTor –SALIENT item at the righthand end of a constituent, where unmarked focus would fall. It is this latter case which has been called anaphoric destressing.

4.3.8 Loose ends

There are a couple of issues which I might be expected to have addressed heretofore but which have not in fact been given more than cursory and/or indirect mention, namely reduction and contraction, and the issue of nuclear versus non-nuclear tones and the nature of contrasts in the post-nuclear region. The following subsections attempt to remedy these omissions somewhat.

4.3.8.1 Kinetic tones versus tonal excursions

In distinguishing between kinetic tones and tonal excursions I was attempting to modify the standard British school approach towards a compatibility with a more Bolinger-like approach. I had felt it a weakness of the British school approach that it could identify only two (or in the case of Halliday, three) significant points in the tone group, namely the head and the nucleus(s). My intuition was that more than one pre-nuclear syllable might have tonal prominence. The British school approach to this has been in some cases to articulate a set of complex head types, and in some cases to put tone group boundaries where I find them unconvincing, in order to preserve the "one kinetic tone per tone group" restriction. Recent proposals at the phonological level by Pierrehumbert [1979, 1980] and Selkirk [1979, forthcoming], which came to my attention after the experiment was run, seem to me to have compelling support, and to represent a more adequate response to the phenomena than the one I proposed, and indeed they are quite similar in spirit.

Briefly, their proposals[15] call for the description of the tone group in terms of an optional initial boundary tone, one or more pitch accents, a phrasal accent, and an optional final boundary tone. The boundary tones and the phrasal accent are specified as either high or low. The pitch accents are either high, low, or a two-tone pair, in which case a specification of which member of the pair is associated with the salient syllable is also required. The height of the highs and depth of the lows is variable, and taken to be controlled by semantic and pragmatic factors. The

[15] In this summary I follow Pierrehumbert rather than Selkirk, although the differences between the two are not great.

last, and possibly only, pitch accent, taken together with the phrasal accent, comprise what has in the past been called the nuclear accent or kinetic tone. Various tone realization, spreading and interpolation principles are included, which determine the actual pitches of each syllable in an utterance given its description in these terms.

The distinction between what I called tonal excursion and the kinetic tone is maintained by this approach, with the extra degree of freedom provided by the phrasal accent accounting for the wider range of tonal activity which distinguishes the kinetic tone. The possibility of the phrasal accent being somewhat distant from the last pitch accent nicely accounts for the perceived unity of such kinetic tones as the *fall ... rise*, which was not uniquely specifiable in my system. However this bipartate nature of the kinetic tone makes the problem of how and where to associate it with the foot structure more difficult. The simple proposal above in section 4.3.7.5 is not adequate to this more complex task, although it will continue to be appropriate for the simpler cases, and some of it will in fact carry over to the more complex ones. The other, and most notable, difference from my original proposal is in the possibility of two-tone, e.g. rising and falling, tonal excursions, an extension which seemed indicated by the experimental results, as mentioned above in section 2.3.

This approach, in common with most previous ones on both sides of the Atlantic, precludes the existence of any tonal activity of any kind after the kinetic tone, beyond a possible steady rise or fall to the boundary. In particular, the modest high tone which typically accompanies salient syllables, which in these terms is a pitch accent with a minimum of oomph, will not be present for post-nuclear feet. My sense is that nonetheless foot structure, being as it is fundamentally rhythmic and not tonal, does exist past the nucleus, albeit without tonal correlate. This has always been the British school view as well. It is fairly difficult to establish beyond doubt, however. There is only one tone group in the data segment which might be germane - *two three times a year I guess.* This was marked with the nucleus on *year* by two subjects, including myself, and with a foot boundary in front of *guess* by two subjects, including myself. This is hardly unequivocal evidence, especially since, as Mark Liberman points out (personal communication), peoples' expectations may be overriding their perceptions. Utterances with significant amounts of post-nuclear material are hard to come by. They rarely occur naturally, and contrived cases usually sound just that. Given Liberman's point, a true test would need to involve a foot boundary not easily predictable from form class alone, for instance one involving the Rhythm

Rule, which makes things even harder. Until someone catches one in the wild and tests a number of subjects on it the point must remain somewhat unclear.

4.3.8.2 Vowel reduction and auxiliary contraction

Vowels in syllables which are –STRESS, or in some cases in non-salient syllables which are +STRESS, are liable to be reduced, that is, to not be fully articulated. Their position of articulation centralizes, producing a schwa, and their duration is lessened. LP's specification of the conditions under which reduction of +STRESS may occur (the *English Destressing Rule*) does not make explicit use of metrical structure, but it does depend on the fundamental wellformedness constraint to block its application in some cases. This suggests that at least the lowest level of metrical structure must be transmitted to the phonological stage, so that the necessary information is available to enforce the constraint. Beyond a second order effect on the timing data presented in Chapter 3, the issue of vowel reduction does not appear to be otherwise relevant to our concerns.

But the question of when the copula and auxiliaries may or may not contract and cliticize onto the preceding word, along with questions about other contractions, would seem to be relevant. I think a plausible account of these phenomena is possible within this framework. Given the amount of attention the issue has received recently, I approach it with some trepidation, especially as I do not have a complete account, but rather am prepared to sketch an answer only to the question of when auxiliaries[16] contract and when they don't. I am indebted to Mark Liberman and Ivan Sag, who first suggested the consequences with respect to this problem of some of the proposals I made above.

The phenomenon under discussion is exemplified by the following:

1) *Who'll bring the beer?*

2a) *John.*
2b) ** John'll.*
2c) *John will.*
2d) *John'll bring it.*

[16] or copular *be* - the approach which follows applies to both but I will refer only to auxiliaries hereafter for brevity.

Simply put, my proposal is that auxiliaries may not contract if they are +FOCUS, and that in (2b) and (2c), as opposed to the others, *will* is +FOCUS, and thus (2b) is blocked. That *will* is +FOCUS follows from the following approach to VP deletion, which follows more or less directly from [Hankamer and Sag 1976] and [Sag forthcoming]. VP deletion is specified at the level of information structure, not grammatical structure. It deletes structure which is -CONTENT as a result of context. Now material which is -CONTENT is necessarily -FOCUS, which implies a marked focus somewhere else. Thus in (2a) we have both the whole verb group and the object noun group deleted, with marked focus on *John*. From this we conclude that both verb group and noun group were -CONTENT. On the other hand, that *will* in (2b) is not deleted altogether implies that the verb group as a whole was *not* -CONTENT, but only *bring* was. Thus within the verb group *will* must have been +FOCUS, since according to the rules given above unmarked focus cannot fall on an element which is -CONTENT. Q.E.D. Note that this implies that *will* is +CONTENT in (2c). That this is the case is supported by the fact that in the proper phonological environment such a +CONTENT auxiliary may start a foot:

3a) */ Gwendolyn / will I'm / sure.*
3b) *? / Gwendolyn will I'm / sure.*

Further evidence for this view comes from the observation that cliticized auxiliaries cannot be highlighted, even when they constitute an independent syllable, while non-contracted ones can. This follows from the requirement that highlighted items must be +FOCUS, while contracted auxiliaries cannot be.

There is one potential problem for this analysis in cases where it can be argued, from a transformational point of view at any rate, that a movement rule, rather than a deletion rule, is responsible for the gap following the potential contraction site. Possible examples of such a situation are in *Ready I am (* I'm) to help you*, from [Lakoff 1970], and *Surprising it is (* it's) that he died so young*. But these cases are highly marked even in their uncontracted form, and not only is contraction blocked, but also there must be a foot boundary before the copula. This explains why no contraction occurs, but it leaves unexplained the origin of the +CONTENT marking on the copula which presumably caused the foot boundary. The answer needs must lie in the exact nature of the information structure in such topic-shifted constructions, but I cannot now present a convincing argument for what that must be. Utterances where the gap is owed to relativization, such as *They saw the place where the statue is (* statue's)*, are intermediate between the clear cut ellipsis cases and these marginal topic-shifted ones. I think the marked focus analysis can be shown to hold for them as well, but more work with naturally occurring data is needed.

Despite these unclear areas, it seems to me that the proposal made above in terms of +FOCUS is a coherent and intuitively plausible approach to the contraction problem in general. It is expressed at the right level, in the right terms, and makes no appeal to traces. Whether the other contraction problems can be solved in a similar matter remains to be seen, although there is some suggestion (e.g. in [Postal and Pullum 1978]) that this is the only real case of contraction, and all the others are cases of lexicalization.

4.3.9 The footmaker: Full formal specification

This chapter is concluded and summarized by the full specification of the footmaker, given on the following pages. It incorporates all the revisions and additions developed in the preceding sections. The italicized section numbers are pointers to the sections where the associated feature or rule is discussed. The first part of **FO** as well as the content, although not the form, of **F2A** and **F2B** are owed to LP. The rest is my responsibility alone.

The Footmaker

Summary of Features

±Stress *4.3.1, 4.3.3.2* – Feature of syllables, determined by phonological properties thereof, as per LP.

±Salient *4.3.2, 4.3.7.1* – Feature of words, determined by the grammatical category thereof, as per the following table:

+Salient: Adjective, Adverb, Demonstrative Pronoun, Intransitive Preposition (e.g. *the / night be / fore*) Noun, Negative Auxiliary (e.g. *that / weren't / opened*), Negative (e.g. *not, never*) , Number, Participle (Past, Passive, Present), Quantifier, Verb, Verbal Particle.

–Salient: Article, Auxiliary, Complementizer, Conjunction,[17] Copula,[18] Determiner, Dummy (*there, it*), Particle (*of* in e.g. */ all of / those, to* in e.g. */ went to / sleep, have to / clean*), Possessive Pronoun, Preposition, Pronoun, Relative Pronoun.

±Content *4.3.6, 4.3.7.3* – Feature of groups or words or rarely syllables (+Content only), assigned prior to the footmaker on a semantic basis, basically to reverse the effects of ±Salient. Propagates from groups to words by following +Focus. Propagates from words to syllables by marking the DTE +Content if the word is +Content, or by marking all syllables -Content if the word is –Content.

+Highlight *4.3.4.2, 4.3.7.4* – Feature of groups or words or rarely syllables, assigned prior to the footmaker on a semantic basis, for emphasis or contrast. Propagates from groups to words by following +Focus, or, in extreme cases, by marking all +Content/+Salient sub-constituents as +Highlight. Propagates from words to syllables by marking the DTE +Highlight if the word is +Highlight.

±Focus *4.3.4, 4.3.7.2* – Feature of intonational sub-constituents, groups, words, or syllables. +Focus is assigned prior to the footmaker on a semantic basis, to indicate the locus of principal communicative content within each constituent. Within the domain of the word, the unmarked location for +Focus is on the DTE of the word. In higher level domains, the unmarked location is on the rightmost sub-element. In the marked case at the word level, the metrical structure of the word is re-arranged to make the +Focus syllable the DTE. In the marked case at higher levels, all sub-elements to the right of that marked +Focus are marked –Focus.

[17] Simple one syllable conjunctions such as *and* or *but* are clearly -Salient. I am not sure about others like *since* or *because* etc.; more data is needed before a sure determination can be made.

[18] This category includes more than just *be* and *become*. Main verbs functioning as copula also are -Salient, as Halliday has pointed out. Compare for instance */ Robin / grew po / tatoes* with */ Robin grew / weary.*

The Footmaker

Summary of Rules

F0 *4.3.1, 4.3.2, 4.3.4, 4.3.7* – Associate with each word in the input constituent a metrical tree and metrical grid as per LP. Propagate feature assignments as necessary. If any non-final immediate constituent of any domain in the input constituent is marked +Focus, mark all immediate constituents to the right of that one –Focus, otherwise, mark the final immediate constituent +Focus.

F1 *4.3.2, 4.3.4.2, 4.3.6* – For all words in the input constituent, normalize and adjust their metrical grid as follows:

1) If the word is +Salient, make the height of the column over the DTE of the word equal to 4. Take the DTE of a one syllable word to be that syllable.

2) Make the height of all columns which are neither over the DTE nor of height 1 equal to 2.

3) For each syllable in the word which is marked –Focus, decrease the height of its column by 1, unless it is already a 1-column.

4) If any syllable of the word is +Content, set the height of its column to 4.

5) For each syllable in the word which is –Content, decrease the height of its column by 1, unless it's already a 1-column.

F2a *4.3.3* – Proceeding from right to left through the metrical grid of the constituent, identify a column as *clashing* on the basis of the membership of the *Clash Set* at that point.

The Clash Set is initially empty (at the right end of the constituent), and its membership is updated at each grid position as follows on the basis of the column at that position:

1) If the column is a 1-column, there is no clash. Remove 2 from the Clash Set and carry on.

2) Otherwise, if the column is a 2-column, then if 2 is a member of the Clash Set, there is a clash. Otherwise, remove 4 from the Clash Set, add 2 to the Clash Set, and carry on.

3) Otherwise, the column is a 4-column. If either 4 or 2 is a member of the Clash Set, there is a clash. Otherwise, add 2 and 4 to the Clash Set and carry on.

F2b *4.3.4* – A *w-s* pair whose DTE is the current syllable may be reversed to *s-w* provided that

1) the new *s* does not immediately dominate a syllable marked –Stress.

2) (the current syllable is not marked +Focus.)

If the current column clashes, there is a strong pressure in favor of performing this reversal if possible.

If the reversal does take place, the latter half of F0 (grid calculation and feature propagation) and F1 are re-applied to the word containing the current syllable, and F2a starts over at the right hand end of the word.

Summary of Rules, continued

F3 *4.3.4.2* – Set the threshold at 2. Proceeding left to right through the constituent grid, at each syllable

1) if the associated column is a 4-column, start a foot with this syllable, set the threshold at 4, and carry on

2) otherwise if the height of the column is greater than or equal to the threshold, and there is a next column whose height is less than the height of this one, either start a foot with this syllable, set the threshold at 4 and carry on, or else decrease the threshold by 1 and carry on. Starting a foot is preferred if the column height is greater than the threshold.

3) otherwise decrease the threshold by one and carry on

F4 *4.3.7.5* – If there is a kinetic tone associated with the input constituent, associate it with the DTE of the word reached by following +Focus down from the top. Pass the results of the whole process on to the next stage

Not all the constructed structure is of course relevant to the next stages. Some information from the previous stages is just passed on, namely specification of tone group boundaries, tonal envelope and boundary tones, kinetic tone type, and word boundaries. New information constructed by the footmaker which is passed on presumably includes foot boundaries, syllable structure and phonemic specification, and the location of tonal accent(s) and kinetic tone.

This distillation of all the discussion which has gone before is quite satisfyingly concise. It satisfies the demands made at the beginning of the chapter, namely that it bring together the extant proposals concerning those factors held to affect foot structure, and that it account for some of the observed variability therein. It uses a small number of independently motivated features to drive a fairly simple mechanism. Some sources of variation are explicit in the rules as stated. The principal source of variation is, however, implicit, deriving from the variation in domain over which the footmaker operates.

The basic structure of the footmaker as specified provides for a clean way of integrating all the various phonological and contextual influences on foot structure. By defining all the features and rules to operate in terms of the metrical grid, and then defining the placement of foot boundaries in terms of the final state of the grid as well, I have avoided a complex multiplicity of specifications of foot structure directly in terms of feature patterns. By using the grid as the common ground, independently specified processes can interact to predict the results of novel coincidences of circumstances. This is of course the basic methodological and esthetic criterion for the adequacy of any framework for expressing linguistic generalizations - namely that it permit the orderly combination of predictions about the effect of features in isolation into predictions about their effect when conjoined.

How well the footmaker and the semantic and pragmatic features it incorporates as here specified meet the demands of the higher goal set out in the introduction, that they serve as the basis for an analysis of discourse structure, will be discussed in the next chapter.

The further question of whether these proposals are realistic, in the sense used back in section 0.1.1, which is to say plausible as the foundation of a psychological model of the relevant aspects of human language production, must be left to the reader. The easy response is to say "That's much too complicated! All those details and numbers and sets and marks - surely I don't have all *those* in my head." But I think such a response takes insufficient account of the complexity of *any* attempt to specify in something approaching an operationally complete fashion any aspect of human linguistic behavior. It's just that to date few such proposals have been made, and we simply aren't familiar with how much is involved. Needless to say, although I perhaps more than anyone else am aware of their defects, I think these proposals are a good beginning on what will be a long and difficult process of extension and refinement.

Chapter 5. Conclusions

In this chapter the formal apparatus of the footmaker is applied to the data segment as transcribed, and a brief overview of what has been accomplished herein is attempted.

5.1 Applying the footmaker to the data segment

In this section, the formal structure proposed in the previous chapter is put to practical use. An informal analysis by synthesis of the prosodic structure of the data segment as transcribed can be performed using the footmaker. Given the assignment of ±SALIENT to its constituent words as prescribed by the footmaker specification, we can consider the question of what additional feature markings are necessary for the elements of each of the tone groups in the data segment for the footmaker as specified to produce the foot structure, tonal excursion, and kinetic tone placement which are found in the data segment, according to the consensus notation of the subjects.

I produced the analysis which follows by hand, in particular not considering all the possible ambiguities of analysis with respect to informational constituent structure. Each tone group is presented, with the foot structure as given in Figure 1 of section 3.1. Above the DTE of each word are listed its feature assignments. Where relevant, the location of tonal excursions is indicated by an up-arrow (↑) and the location of kinetic tones is indicated by a bullet (•). The basis for these markings, that is, whether they were unanimous, mine alone, or whatever, is given in the subsequent discussion, which will also point out for each tone group interesting and/or problematical aspects of the analysis. Square brackets are used where necessary to indicate the domain of ±FOCUS marking, and a small circle (o) indicates the locations of pauses, filled or unfilled.

```
-SAL     +SAL  -SAL     -SAL  -SAL    +SAL         +SAL  -SAL   -SAL  -SAL    +SAL  -SAL
1SPRO    V     COMP     PREP  ART     N            IPREP DUM    COP   ART     Q     PART
I     /  know  that  o  on    the  /  night o be / fore  there  was   a    /  lot   of   o ek spec

       +SAL       -SAL        +SAL        -SAL  -SAL       +SAL
       N          CONJ        N           PREP  POSPRO     N
    /  ta   tion  nd   ek  /  cite  ment  on    my      /  part  |
```

Some subjects put foot boundaries in front of the pronouns *I* and *my*. This would imply +CONTENT markings on those words

```
               + CONT
 -SAL          -SAL    -SAL  -SAL    -SAL    + SAL
 CONJ          1SPRO   AUX   PART    PART    V
/ af   ter /   I       went  to  o   to  /   sleep  |
```

Here we have a disyllabic preposition at the beginning of a tone group getting a foot boundary, as discussed in sections 4.1.3 and 4.3.5. Uniform agreement on the foot boundary before *I* prompts the +CONTENT marking there.

```
-SAL     + SAL           -SAL       -SAL      + SAL          + SAL          + SAL     + SAL         -SAL  -SAL
POSPRO   N               AUX        AUX       ADV            V              VPART     Q             PART  ART
my    /  par   ents o    would o    would /   al   ways /    o    pen /     up   /    se    vrel    of    the

       + SAL
       N
     / gifts  |
```

The unmoved verbal particle *up* is classed as +SALIENT

```
                                                       + HIGH ↑    ●
                                                       + FOC      -FOC
  + SAL  -SAL  -SAL        + SAL        + SAL  -SAL    + SAL      + SAL
  Q      PART  ART         ADJ          Q      ART     ADJ        N
/ se     vro   the  im /   por  nt /    all    the [ su / prise   things  ] |
```

Here we see a slight problem. The marked focus accompanying highlighting has bleached the following foot boundary, although the fact that one subject (me) marked a boundary there indicates some uncertainty. The real problem for the system lies in the location of the kinetic tone, which was marked by four out of five subjects as shown. This is in fact an example of the kind of situation where the Pierrehumbert scheme is more appropriate. As noted above in 4.3.8.1, my system as specified has no way of assigning the 'dispersed' kinetic tones to the right place, as the placement is specified as adhering to +FOCUS. In this case we have the pitch accent component of the nuclear tone adhering to +FOCUS all right, but the phrasal accent is further along. Clearly more work is needed to specify what patterns of tune-text association are possible, although I expect that the association of the nuclear pitch accent with +FOCUS will stand up.

```
                                                        -CONT          + CONT
-SAL     -SAL  -SAL      + SAL             + SAL         -SAL   + SAL          -SAL      -SAL    + SAL
CONJ     CONJ  PREP      N                 N             2SPRO  V              1SPRO     AUX     V
and  /   so    on   /    christ  mas  /    mor   ning    you    know      /    I         would   wake

       + SAL
       VPART
     / up   |
```

So is one of those conjunctions whose behavior suggests that calling all conjunctions -SALIENT is an oversimplification. The verb *know* is marked -CONTENT in the parenthetical context. The verb-particle construction presents a problem here. Perhaps the right solution would be to allow marked focus on the rightmost element in a group, as distinguished from unmarked focus, to assign -FOCUS to the elements to its left, which would bleach the preceding foot boundary as in this case, but leave the one in front of *open* in the last tone group but one possible, owing to the disyllabic nature of the verb.

```
                                      •
                                    +Foc     -Foc
-Sal    +Sal       -Sal    -Sal     +Sal     +Sal
AUX     V          PREP    ART      PRPRT    N
go  /   run  ing   in  to  the  / [ liv  ing room  ] |
```

The figure/ground viewpoint on focus in noun phrases seems quite apt to expressions such as *living room.* The marked focus carries the kinetic tone and bleaches the subsequent foot boundary as predicted, although not all agreed on the kinetic tone location.

```
        +Cont                                                                   +High ↑
-Sal    -Sal    -Sal      -Sal    +Sal    -Sal     +Sal   -Sal     +Sal          +Sal
CONJ    3SPRO   AUX       COP     PSVPRT  PREP     Q      DET      ADJ           N
and /   it      would o   be /    filled  with /   all    these /  won  drous /  things  |
```

There does not seem to be any grounds for a marked focus on *things* here.

```
                  +Cont
-Sal   -Sal       -Sal      +Sal      +Sal             -Sal      +Sal       +Sal
DUM    AUX        COP       Q         N                RELPRO    NAUX       V
there  would /    be     /  some  /   pre  sents o     that   /  werent  /  o    pened  |
```

The copula is sometimes +Content, especially in existential contexts

```
                                                                  +Cont
  +Sal           -Sal     -Sal     +Sal              -Sal     -Sal     -Sal     -Sal        -Sal     +Sal
  ADV            ART      ART      N                 RELPRO   AUX      COP      PREP        ART      N
/ ty    pi gli   the  o   the  /   pre   sents       that     had  /   been     un   der    the  /   tree  o  be

      +Sal      +Sal
      IPREP     ADV
    / fore  /   a    ny way  |
```

```
-Sal     -Sal       +Sal            +Sal   -Sal       +Sal
1PPRO    AUX        V               Q      PART       DEMPRO
we       would /    o · pen  /      all    of    /    those   |
```

Note that demonstrative pronouns are +Salient, in contrast to demonstrative determiners.

```
   +High ↑
   +Foc      -Foc           -Cont             -Cont
   -Sal      +Sal           +Sal      +Sal    +Sal    -Sal     +Sal          -Sal    +Sal          +Sal
   1SPRO     V              ADV       PRPRT   ADV     ART      ADV           ART     ADV           ADJ
/ [ I        re  mem  ber ] just   /  bing    just    a    /   ve   ry o     a   /   re   ly  /    su    per

                                         +High ↑
      +Sal          +Sal    -Sal         +Sal
      ADJ           Q       PART         N
    / ha   py  /    kin     of    /      time     |
```

The focusing here was discussed above in section 4.3.7.2. There is some suggestion that *just* is actually -Salient. The marked case for it seems to be when it *does* begin a foot, not as here when it fails to.

```
                                 +FOC    -FOC                                           +FOC        -FOC
-SAL   -SAL    +SAL    +SAL      +SAL    +SAL    +SAL   -SAL      -SAL   -SAL           +SAL        +SAL
1PPRO  AUX     V       N         PSVPRT  VPART   ADV    PREP      ART    ART            PRPRT       N
we     would / leave / things / [spread  out   ] all  /  o   ver  the o  the /[  liv   ing  room ]

       ●
       +SAL
       N
     / floor  |
```

The verb-particle construction here is focused opposite to those above. *All over* is one of the adverb-preposition pairs with a single metrical structure to which Iambic Reversal may apply, although it has not done so here. The individual salience markings are irrelevant here - the whole three-syllable collocation should be marked +SALIENT. Given the location of the kinetic tone, we assume an unmarked focus at the top level of the complex noun phrase which ends the clause.

```
  +SAL     -SAL  -SAL    +SAL
  N        PREP  ART     N
/ things   in    a     / mess  |
```

```
       -CONT           +HIGH ↑                   +HIGH ↑
-SAL   +SAL   -SAL     +SAL      -SAL    -SAL    +SAL    +SAL    -SAL       -SAL    +SAL
2SPRO  V      PART     Q         ART     PART    Q       Q       PREP       DET     N
you    know   at    /  least  o  the  o  at   /  least   all  /  through    that /  day    |
```

No focus is indicated here in conjunction with the highlighting, as the constituent structure is not clear. Iambic Reversal has once again not applied to the adverb-preposition pair here, as discussed in section 4.3.5.

```
  +SAL            +SAL         +SAL
  ADV             Q            N
/ some  times /   se    vrel / days  |
```

```
  +SAL     -SAL    +SAL  -SAL     +SAL
  DEMPRO   COP     Q     PART     ADJ
/ that     was /   kind  of   /   spe   cial  |
```

```
                                       +CONT
-SAL      +SAL   -SAL  -SAL     +SAL     +SAL        +SAL
1PPRO     NAUX   AUX   PART     V        N           VPART
we    /   din    ave   to    /  clean o  things  /   up     |
```

It is not clear to me whether the lack of a foot boundary for *things* is caused purely locally, as indicated here, or is in fact tied up with particle movement - more data is needed.

```
-SAL   -SAL    +SAL
3SPRO  COP     ADJ
it     was /   good  |
```

```
        +HIGH ↑
-SAL   +SAL     -SAL  +SAL  -SAL   +SAL  -SAL    +SAL   +SAL       -SAL -SAL  +SAL
ART    ADV      ART   N     1SPRO  V     RELPRO  V      ADJ        PART ART   N
the  / most  o  the / thing  I   / think that  / comes/ clos est o to   a   / fam ly

        +SAL
         N
     /  ri   tu al |
```

```
        +CONT
-SAL    -SAL      +SAL
PREP    POSPRO     N
in  /   my    /  fam  ly |
```

```
  +CONT
  -SAL     +SAL      -SAL    -SAL       +SAL
  COP       N        PREP    POSPRO      N
/ was   /   vi  zits  to  o   my    / grand par ents |
```

The +CONTENT marking on *was* here seems to result from the long pauses which separated it from the beginning of its clause, setting it up as a semi-independent predication.

```
                      +FOC   -FOC
-SAL       +SAL       +SAL   +SAL           +SAL       +SAL
RELPRO      V         NUM     N              N         IPREP
who    /  lived  / [ four    hun    red ] / miles a /  way    |
```

Marked focus in a measure phrase.

```
      -CONT
-SAL  +SAL      +SAL
PREP  ADJ        N
in    new    / mek  si  ko |
```

An alternative here would be to call *New Mexico* a single word with one metrical structure.

```
                                                                   -CONT
  +SAL       +SAL     -SAL   +SAL        +SAL  -SAL     +SAL       +SAL      +SAL
  DEMPRO     ADV      ART     Q           N    3SPRO    ADV         V        ADV
/ thats  /   re   ly  the /   on   ly / trips   we  /    e    ver  took  /    a    ny where |
```

As discussed above in section 4.3.6, *took* is contextually redundant.

```
      -CONT                                   -CONT
-SAL  +SAL    -SAL      +SAL       -SAL   +SAL  +SAL        +SAL
1SPRO  V      POSPRO     N          AUX   NEG    V           N
 I    mean     my   /    fa   mli  did /  not   take   va /  ca   tions |
```

Two -CONTENT verbs here, one parenthetical, the other redundant.

```
-SAL  -SAL   -SAL     +SAL    +SAL       +SAL      -SAL    +SAL       +SAL
CONJ  1PPRO  AUX       V       V          N        CONJ     N          N
but    we    would /   go  /   vi   sit / gran ma   n  /  gran pa /  reed   |
```

+SAL	+SAL	+SAL	-SAL	+SAL	-SAL	+SAL
NUM	NUM	N	PART	N	1SPRO	V
/ two	/ three	/ times	a	/ year	I	/ guess |

We don't get a measure phrase marked focus here - possibly because of the list.

			+FOC	-FOC				
-SAL	-SAL	-SAL	+SAL	+SAL		+SAL		+SAL
DUM	COP	DET	NUM	N		N		N
there	was	that	/ [eight	ow	er]	/ au	to	/ trip |

Measure phrase marked focus.

-SAL	+SAL	-SAL	+SAL	+SAL		+SAL
1SPRO	V	ART	N	ADV		ADV
I	/ knew	the	/ road	/ ve	ry	/ well |

+CONT			-CONT			
-SAL	+SAL	-SAL	+SAL	+SAL		
3PPRO	V	PREP	ADJ	N		
/ they	/ lived	in	new	/ mek	si	co |

Unpredictable change of subject/theme, so subject pronoun is +CONTENT.

				-CONT		-CONT										
-SAL	+SAL			+SAL	-SAL	+SAL	-SAL	+SAL		+SAL			+SAL		+SAL	
PREP	ADV			Q	PART	Q	PART	ADV		ADJ			ADJ		N	
in	/ ve	ry	o	ki	na	sort	of	/ se	mi	/ a	re	a	/ a	rid	/ coun	try |

Those -CONTENT quantifiers were uttered very fast, and all in all lack punch.

+SAL	
ADJ	
/ san	dy |

	-SAL	-SAL	-SAL	+SAL		+SAL		+SAL	-SAL	+SAL	+SAL
	CONJ	DUM	COP	ADV		ADJ		Q	PART	N	ADV
o	so	it	za	/ ve	ry	/ dif	rent	/ kin	of	/ world	/ there |

+CONT								
-SAL	+SAL		+SAL		-SAL	+SAL		
3PPRO	ADV		V		1SPRO	ADV		
/ they	/ al	ways	/ trea	ded	me	/ won	der	fly |

Again, unpredictable subject pronoun.

									•	
								+FOC	-FOC	
-SAL	+SAL			+SAL	-SAL	-SAL	-SAL	+SAL	+SAL	
POSPRO	N			V	PREP	PREP	ART	N	N	
my	/ grand	fa	ther	/ work	tout	in	the	/ [oil	fields	] |

Not everyone agreed on the location of the kinetic tone here.

-SAL	-SAL	-SAL	+SAL		+SAL	-SAL	+SAL		
3SPRO	AUX	AUX	ADJ		Q	PART	N		
he	was	a	/ var	ious	/ kin	of	/ sales	man	|

-SAL	+SAL		+SAL	
PREP	ADJ		N	
at	/ diff	rent	/ times	|

+CONT				+CONT		
-SAL	-SAL	+SAL	-SAL	-SAL	-SAL	
3SPRO	AUX	V	1SOPRO	PREP	3SOPRO	
/ he	would	/ take	me	/ with	him	|

No good explanation for the +CONTENT on *he* here, although there was a long pause before this tone group, and so in some sense represents the speaker's active choice to continue with this theme rather than switching.

-SAL	-SAL	+SAL	
CONJ	3SPRO	V	
when	he	/ went	|

-SAL	-SAL	
CONJ	CONJ	
and	/ so	|

So again.

+CONT				
-SAL	-SAL	+SAL		
3PPRO	AUX	V		
/ we	would	/ tra	vel	|

		•	
+CONT	-CONT	-CONT	
-SAL	+SAL	+SAL	
1SPRO	NAUX	V	
/ I	do	know	|

Weird little parenthetical tone group. I think this is again a case of an extended kinetic tone covering the whole of the foot, with the phrasal accent on the last word. Some subjects did mark a foot boundary before *know*.

+FOC	-FOC								
+SAL	+SAL		+SAL	-SAL	+SAL	-SAL	+SAL		
NUM	N		N	ART	N	CONJ	N		
/ [two	hun	dred]	/ miles	a	/ day	or	/ some	thing	|

```
-Sal    -Sal       +Sal            +Sal
3SPRO   AUX        V               IPREP
 he    would /   tra   vel a /   round   |

   +Sal   -Sal     +Sal
 DEMPRO  CONJ    DEMPRO
/  this    and  /   that    |
```

5.2 What does it all mean, and where do we go from here?

Whether or not the footmaker represents a valid model of either competence or performance in the prosodic domain cannot of course be determined by its application to 130 seconds of speech. I hope the preceding exercise has demonstrated its plausibility for the task, and that in fact the way is now prepared for future work to start from this base and actually proceed to investigate discourse structure using natural spoken data. I think most if not all the feature marking given above for the data segment as necessary to generate the foot structure as notated is in fact semantically plausible. Furthermore, my post-hoc interpretation of the text is consistent with those markings, and does not suggest any others.

Despite the difficulties with some aspects of the transcription system presented in chapter 2, I think in fact few modifications are required to make it do as good a job as is possible without instrumental assistance. What I mean by this is that the weakest area of the system was tune identification, along with confusions with respect to the distinction between kinetic tones and tonal excursions. But the work of Pierrehumbert and Selkirk (op. cit.) leads inescapably to the conclusion that mechanical assistance in the form of pitch tracking and display is an essential prerequisite to the accurate notation of pitch movement. There are too many degrees of freedom, and the temporal discriminations required are too fine, for even the best trained ear to be able to reliably achieve an accurate "tonemic" transcription. But all is not lost nonetheless.

People *are* reliable about positional facts - tone group and foot boundaries, and kinetic tone location. I think that with some modification of the system along the lines suggested in section 2.3 to improve the consistency of discrimination between marked pitch accent (my tonal excursion) and the modest pitch accent accompanying foot boundaries, that the *location* of all prosodic phenomena could be consistently transcribed with a small amount of training. And this locational information in fact provides almost all of the information necessary for discourse analysis. The footmaker as described makes no use of tune type, and the informational features it is concerned with can be determined without reference thereto. I think it can be argued that these features or some like them are in fact all that are needed to account for almost all the recognized aspects of intonational meaning, as has been argued in [Liberman and Thompson 1980]. In any case the best way to test this hypothesis is to attempt to use this system in the analysis of richer and more extensive texts. If it provides the basis for convincing insights into the structure of such texts, then I will be well satisfied, for that is the best validation of a theory which can be hoped for.

I hope to see the system used for production as well as analysis. Progress proceeds apace at the lower levels of synthesis by rule systems, and I think the time is right to attempt to incorporate structure above the word level into such systems, and indeed some attempts in this direction have already been made. Given the explicitly production oriented nature of the footmaker specification, it should be possible to incorporate it into a synthesis system without too much difficulty, and I hope this will happen.

In conclusion I hope that beyond the contributions to the specific problem areas addressed in this thesis, three larger points have emerged from these pages: Live data is the ultimate foundation on which all our linguistic speculations rest, and it is a good idea to constantly refer those speculations back to the data; No one individual's perceptions are so accurate that they cannot stand to be backed up by experimental support in terms of controlled experiments and carefully presented, easily comparable, statistics reporting the results thereof; Processing models of linguistic behavior are a legitimate vehicle for expressing theoretical insights, and can provide a supportive framework or paradigm for the pursuit of such insights.

Appendix A. Glossary

I use the following abbreviations in the definitions which follow: FP - functional perspective, PhP - phenomenal perspective, FC - functional correlate, PhC - phenomenal correlate. The numbers in italic indicate the section where these terms are first discussed.

accent: PhP, following Bolinger, I reserve this for what actually occurs in utterance tokens. Roughly equivalent to my *tonal excursion 1.1.2.1.3*

amplitude: The amplitude of the acoustic speech signal - the principal determinant of perceived *loudness.* Often not distinguished from intensity or loudness.

anaphoric destressing: What happens to full noun phrases like *the turkey, the bastard, the sweetheart* when they are used more or less as pronouns, as in the difference between *This guy slammed on his brakes in the middle of nowhere, so of course I ran into the turkey* and *I came around the corner and there were a bunch of chickens and a turkey. The chickens scattered, but I ran into the turkey.* Ladd [1978] says that the perceived prominence of the verb results from the *anaphoric destressing* of the following NP. I propose a different account in chapter 4. *4.3.7.5*

boundary tone: Theoretical construct used to describe aspects of *intonational words* which seem to occur at the first and last syllable of *tone groups,* whether salient or not. *1.1.2.2.1*

categorial: FP, sub-system of *prosody,* FC of *tonal* acoustic aspects of utterances. *1.1.2*

Clash Detection: Constituent part of my formulation of the Rhythm Rule. *4.3.3.1*

clash set: The formal expression within my formulation of Clash Detection of metrical grid configurations which would clash to the left of a given location therein. *4.3.3.1*

contour: Same as *kinetic tone.* Sometimes also used as roughly synonymous with *intonational word,* as in the *declarative contour, the surprise-redundancy contour. 1.1.2.2.1*

dactyl: A foot consisting of three syllables *1.1.2.1.4*

DTE: Short for Designated Terminal Element. The DTE of a node in a metrical tree is that terminal of the tree reachable from that node by always following the strong branch downwards - that is, the strongest terminal dominated by the node. *4.3.1.2*

duration: The temporal extent of segments of the acoustic speech signal - the principal determinant of perceived *length*. Often not distinguished from length.

envelope: Theoretical construct used to describe the aspect of *intonational words* which establishes the upper and lower limits of pitch movement within a *tone group*, thereby also determining what counts as high or low pitch. *1.1.2.2.1*

foot: Halliday's word, following Abercrombie, for the basic unit of the rhythmic structure of English. For him, this is the fundamental percept, and *salience* is a derivative notion. See also *stress*. *1.1.2.1.3*

f0: Pronounced *ef zero* or *ef oh*. Abbreviation for *fundamental frequency*.

fundamental frequency: The fundamental frequency of the acoustic speech signal - the principal determinant of perceived *pitch*. Often not distinguished from pitch.

highlighting: FP, my name for the situation in which a syllable, and possibly by extension its parent word and constituent, is made to stand out, FC of *tonal excursion*. *1.1.2.2.2*

Iambic Reversal: LP's name for their formulation of the shift in metrical structure which accomplishes the retraction in the Rhythm Rule. Also a constituent part of my formulation of the Rhythm Rule. *4.3.3.2*

information focus: The element of any constituent in information structure which bears the principal functional load of that constituent is said to be focused. This is my extension of Halliday's concept. *4.1.6, 4.3.4*

information structure: FP, Halliday's concept of a structuring of the utterance independent of, although often paralleling, its syntactic structure, based on its status as information. *4.3.7.2*

information unit: FP, Halliday's name for FC of the *tone group*. *1.1.2.1.2*

intensity: Strictly speaking (following Crystal) the property of the articulation of speech which conditions *amplitude*. Often not distinguished from amplitude or loudness.

intonation: Cover term for all phenomenon which are pitch related. Sometimes not distinguished from *prosody*. Most narrowly, as in when comparing the *intonation system* with the *salience system*, restricted to phenomena associated with the *intonational word*, the *tone group*, and their relation.

intonational word: Specifies both an *envelope*, *boundary tones*, and *kinetic tone* (PhP), on the one hand, and some sort of meaning schema (FP) on the other. The surprise-redundancy contour of Liberman and Sag [1974] is an example. The basic idea goes back at least as far as Trager and Smith [1951], and extends forward from there. Ladd [1978] gives the most extensive discussion of this to date. *1.1.2.2.1*

kinetic tone: A perceived movement of pitch, usually either within a single (salient) syllable or across a disyllable. *1.1.2.2.1*

length: The perceptual property of speech regarding the extent in time of e.g. segments. Principally determined by *duration,* from which it is often not distinguished.

LP: Short for Liberman and Prince [1977]. *4.3.1*

loudness: The perceptual property of speech regarding the strength of the auditory impact of e.g. segments. Principally determined by *amplitude,* from which it is often not distinguished.

metrical grid: LP's formal structure for relating stress and foot structure and tune. It is positional and absolute, in contrast with metrical structure (q.v.), which is tree-structured and relational.

metrical structure: LP's tree structured, relational notation for stress. It applies at the word level, and possibly at the constituent level as well. It consists of a binary tree, with each node representing a weak-strong relation between its two descendants. At the lexical and compound level, it captures all the same phenomena as the *CSR* and *ESR* (see *stress)* and then some, in a very pleasing way. Arguments for its utility above the level of compounds are less clear. *See anaphoric destressing, Rhythm Rule,* and *tune-text association.*

The realization of this abstract stress structure as salience is accomplished by alignment with the *metrical grid* (q.v.). There is some fuzziness about whether the boundary between stress and salience is kept clear in cases where they are not identical. See *anaphoric destressing, Rhythm Rule.*

msf: A foot consisting of only one syllable. *1.1.2.1.4*

organizational: FP, sub-system of *prosody,* FC of *temporal* acoustic aspects of utterances. *1.1.2*

pitch: The perceptual property of speech regarding the position on a scale from low to high of e.g. segments. Principally determined by *fundamental frequency,* from which it is often not distinguished

prosody: Non-segmental acoustic aspect of utterance. *1.1.1*

qsf: A foot consisting of four syllables. *1.1.2.1.4*

Rhythm Rule: The name given to the retraction of foot boundaries in the presense of a subsequent rhythmic clash, as in */ Tennessee / border* as opposed to *Tennes / see. 4.1.2, 4.3.3*

RPPR: Short for Relative Prominence Projection Rule. This is LP's statement of the relation between the metrical tree and the metrical grid. *4.3.1.2*

salient: PhP, that property of the first syllable of a foot which distinguishes it from the rest. That is to say, the first syllable of each *foot* is salient, the others if any are non-salient. It is important to note that this is a derivative notion - the foot is fundamental. *1.1.2.1.3*

stress: Following Bolinger, the locus of potential *salience 1.1.2.1.3*

temporal: PhP, sub-system of *prosody*, PhC of *organizational* structure, covers the durational and rhythmic aspects of acoustic form. *1.1.2*

thirteen men rule: Alternative name for the Rhythm Rule, which it exemplifies.

tonal: PhP, sub-system of *prosody*, PhC of *categorial* structure, covers the fundamental frequency and amplitude aspects of acoustic form. *1.1.2*

tonal excursion: PhP, a short period of high or low pitch which is significantly different from the surrounding pitch, PhC of *highlighting. 1.1.2.2.2*

tonality: Halliday's word for the system which controls the division of speech into tone groups.

tone: Halliday's word for the system which controls the choice of kinetic tone.

tone group: PhP, rhythmically and tonally coherent stretch of speech. FC is the *information unit. 1.1.2.1.1*

tonicity: Halliday's word for the system which controls the location of the kinetic tone within the tone group.

trochee: A foot consisting of two syllables. *1.1.2.1.4*

tune: Pretty much the same as *kinetic tone* - can be used of any distinctive pattern of pitch movement, large or small in scope. *1.1.2.2.1*

tune-text association: Liberman's name (in [Liberman 1975]) for the problem of determining the relation between the segmental and intonational structures of an utterance. *1.1.2.2.1*

Appendix B. Sample data

The next two pages reproduce the transcription worksheets for subject C's fifth trial on both the foot boundary task and the tone task.

Ex # 1a with confidence rating

Use this form for your transcription. Please fill in your name, the trial number, the date, and the time at which you start. Try to do the trial at one sitting - if you are interrupted record the time of the interruption in the space provided at the end.

Your name: CC
Trial #: 5
Date of trial: 10/16/79
Start time: 8:10 pm - 8:25 pm

/I know that um on the/night be-/fore/there was a/lot of uh ex-pec/ta-tion and ex-/cite-ment on /my/part and um af-ter/I went to to/sleep my/pa-rents/would would/al-ways/o-pen/up um /se-ve-ral of the/gifts/se-ve-ral of the im-/por-tant/all the sur-/prise things and/so on/christ-mas /morn-ing/you know/I would wake/up and go/run-ning in-to the/li-ving room and/it would be /filled with/all these/won-drous/things um and/then/there would be/some/pre-sents that/weren't /o-pened/ty-pi-cal-ly the the/pre-sents that/had/been/un-der the/tree be-/fore/a-ny-way and/we would/o-pen/all of/those and/I re-/mem-ber just/be-ing just a v a/ve-ry a v a/real-ly/su-per /hap-py/kind of/time and/we would/leave things spread/out all o-ver the k the/li-ving room/floor and/things in a/mess for/you know at/least the en at/least all/through that/day/some-times /se-ve-ral days and/that was/kind of/spe-cial we/did-n't have to/clean/things/up/and it was/good the/most the/thing I/think that/comes/clo-sest to a/fam-i-ly/ri-tu-al in/my fam-i-ly was uh/vi-sits to um my/grand-par-ents who/lived um/four hun-dred/miles a-/way in um new/me-xi-co and /that's/real-ly the/only/trips we/e-ver took/a-ny-where I mean my/fam-i-ly did/not take va-/ca-tions but/we would/go vi-sit/gran-ma and/gran-pa/reed um/two/three times a/year I/guess so/there was that/eight ho-ur/au-to trip I/knew the/road ve-ry/well and uh they lived in new/me-xi-co in/ve-ry uh/sort of/se-mi/ar-e-a/a-rid/coun-try/san-dy and/so it was a/ve-ry/dif-fe-rent kind of/world there and they/al-ways/trea-ted me/won-der-ful-ly my/grand-fa-ther/worked out in the/oil fields /he was a/va-ri-ous/kind of/sales-man at/diff-er-ent/times and/he would/take me/with him when he/went and/so we would/tra-vel/I don't know/two hun-dred/miles a/day or/some-thing he would /tra-vel a-/round/this and/that

Interruption start:
Interruption end:

End time:

Thank You

Ex # 1b with confidence rating

Use this form for your transcription. Please fill in your name, the trial number, the date, and the time at which you start. Try to do the trial at one sitting - if you are interrupted record the time of the interruption in the space provided at the end.

Your name: CC
Trial #: 5
Date of trial: 10/16/79
Start time: 8:33 pm - 8:45 pm

I know that um on the night be–fore | there was a lot of uh ex–pec–ta–tion and ex–cite–ment on
my part | and um af–ter I went to to sleep | my pa–rents would would al–ways o–pen up um
se–ve–ral of the gifts | se–ve–ral of the im–por–tant all the sur–prise things | and so on christ–mas
morn–ing you know I would wake up | and go run–ning in–to the li–ving room | and it would be
filled with all these won–drous things | um and then there would be some pre–sents that weren't
o–pened | ty–pi–cal–ly the the pre–sents that had been un–der the tree be–fore a–ny–way | and we
would o–pen all of those | and I re–mem–ber just be–ing just a v a ve–ry a v a real–ly su–per
hap–py kind of time | and we would leave things spread out all o–ver the k the li–ving room floor |
and things in a mess | for you know at least the en at least all through that day | some–times
se–ve–ral days | and that was kind of spe–cial | we did–n't have to clean things up | and it was good |
the most | the thing I think that comes clo–sest to a fam–i–ly ri–tu–al | in my fam–i–ly | was uh vi–sits
to um my grand–par–ents | who lived um four hun–dred miles a–way | in um new me–xi–co | and
that's real–ly the only trips we e–ver took a–ny–where | I mean my fam–i–ly did not take va–ca–tions |
(maybe just) | → but we would go vi–sit gran–ma and gran–pa reed | um two three times a year I guess | so there was
that eight ho–ur au–to trip | I knew the road ve–ry well | and uh they lived in new me–xi–co | in ve–ry
uh sort of se–mi ar–e–a a–rid coun–try | san–dy | and so it was a ve–ry dif–fe–rent kind of world
there | and they al–ways trea–ted me won–der–ful–ly | my grand–fa–ther worked out in the oil fields |
he was a va–ri–ous kind of sales–man | at diff–er–ent times | and he would take me with him when
he went | and so we would tra–vel I don't know two hun–dred miles a day or some–thing | he would
tra–vel a–round this and that

Interruption start:
Interruption end:

End time:

Thank You

Appendix C. Support systems

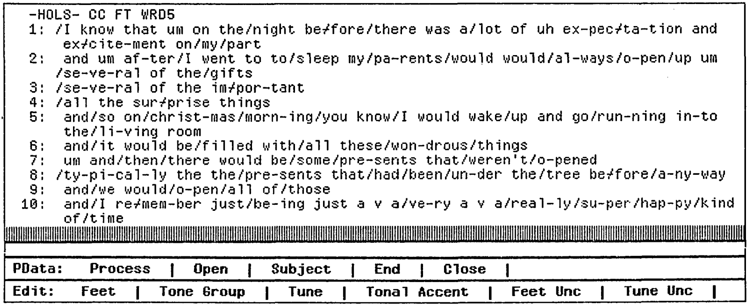
This appendix contains brief descriptions of the three computer systems which were used in support of various aspects of the thesis.

C.1 Data sheet transcription

Preparing for statistical analysis twenty-nine. data sheets for each of two tasks, each with over a hundred marks of one sort or another, represented a task of significant proportions. I developed an extensive computer system to make the task as fast, accurate and painless as possible. The system was written in Interlisp (see [Teitelman 1978]), and made use of the high resolution graphic display input and output facilities available within the Interlisp environment at Xerox PARC (described in [Sproull 1979], [Thompson 1978]). These facilities allowed me to write programs which displayed the text of the data segment on a screen, and then indicate thereon with a hand-held pointing device the location and nature of the marks made by a particular subject on a particular trial. On completion, these displays could then be converted to numerical form for input to the statistical system described in the next section. The two figures below show snapshots of the display screen as it appeared when the system was being used to re-transcribe the worksheets shown in the preceding appendix. Figure 1 is part way through the re-transcription of the foot boundaries, Figure 2 part way through the re-transcription of the kinetic tones.

The top region of the picture shows the text being processed. The first line identifies the text - HOLS, the subject - CC, the type of data being transcribed - FT (for foot boundaries) or TN (for tune), and the number of the trial - 5. The two rows of boxes at the bottom of the picture are "menus" for controlling the system. By pointing at one of them with the pointing device and pushing a button on it, the indicated operation can be invoked. The bottom menu controls the type of information being transcribed - feet in Figure 1, and tune in Figure 2. The foot boundaries in Figure 1 were inserted into the picture by pointing with the pointing device and pushing the button on it, moving on to the next, and so on. There are actually three buttons on the device, so another button can be used to remove misplaced marks. Figure 2 shows the kinetic tone marking, which makes use of combinations of buttons to indicate falling (\ ), rising (/ , not shown in this example), falling-rising (V), and rising-falling (A), inserted into the picture just before the syllables they modify. The striped band at the bottom of the text region indicates that more text remains and can be displayed by pointing and buttoning in that area.

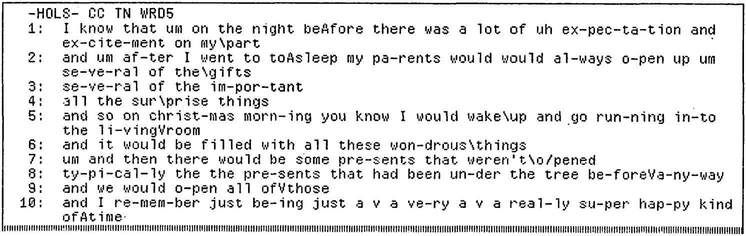
This system proved extremely helpful, and enabled me to enter the data for a trial - six sorts of marks from the two sheets - in about half an hour, including double-checking.



```
  -HOLS- CC FT WRD5
  1: /I know that um on the/night be⌿fore/there was a/lot of uh ex-pec⌿ta-tion and
       ex⌿cite-ment on/my/part
  2:  and um af-ter/I went to to/sleep my/pa-rents/would would/al-ways/o-pen/up um
      /se-ve-ral of the/gifts
  3: /se-ve-ral of the im⌿por-tant
  4: /all the sur⌿prise things
  5:  and/so on/christ-mas/morn-ing/you know/I would wake/up and go/run-ning in-to
       the/li-ving room
  6:  and/it would be/filled with/all these/won-drous/things
  7:  um and/then/there would be/some/pre-sents that/weren't/o-pened
  8: /ty-pi-cal-ly the the/pre-sents that/had/been/un-der the/tree be⌿fore/a-ny-way
  9:  and/we would/o-pen/all of/those
 10:  and/I re⌿mem-ber just/be-ing just a v a/ve-ry a v a/real-ly/su-per/hap-py/kind
       of/time
```

PData: Process | Open | Subject | End | Close |
Edit: Feet | Tone Group | Tune | Tonal Accent | Feet Unc | Tune Unc |

Figure 1. Snapshot of screen while re-transcribing foot boundaries.



```
  -HOLS- CC TN WRD5
  1:  I know that um on the night beAfore there was a lot of uh ex-pec-ta-tion and
      ex-cite-ment on my\part
  2:  and um af-ter I went to toAsleep my pa-rents would would al-ways o-pen up um
      se-ve-ral of the\gifts
  3:  se-ve-ral of the im-por-tant
  4:  all the sur\prise things
  5:  and so on christ-mas morn-ing you know I would wake\up and go run-ning in-to
      the li-vingVroom
  6:  and it would be filled with all these won-drous\things
  7:  um and then there would be some pre-sents that weren't\o/pened
  8:  ty-pi-cal-ly the the pre-sents that had been un-der the tree be-foreVa-ny-way
  9:  and we would o-pen all ofVthose
 10:  and I re-mem-ber just be-ing just a v a ve-ry a v a real-ly su-per hap-py kind
      ofAtime
```

Figure 2. Snapshot of screen while re-transcribing kinetic tones.

C.2 IDL - the Interactive Data-analysis Language

This section briefly describes the interactive statistical system I used to compute all the measures presented in chapters 2 and 3. The system is described in detail in the IDL reference manual [Kaplan, Sheil, and Smith 1978], to which the interested reader is referred for more detail. A few quotes from the introduction to this manual give an idea of the philosophy of the system.

> "The Interactive Data-analysis Language (IDL) is a programming language designed for the analysis of social science data. IDL is radically different from most previous computer systems for social science data analysis, as it is not based on the statistical "package" or subroutine concept, but on the concepts employed in the design of modern programming languages. Thus, IDL provides the user with powerful tools for the manipulation of the type of data most commonly used in social science research, rectangular numeric arrays, and the ability to combine these tools to yield a wide range of statistical analyses. The sophisticated computer user will find similarities between IDL and Iverson's APL, at least in the facilities for handling arrays, but IDL goes well beyond APL in providing capabilities especially designed for practical data analysis, such as the handling of labels (including the automatic generation of new labels by system routines) and missing data, and the ability to analyze very large data sets." ...

> "The approach being followed by IDL is to give the user a powerful set of statistical "building blocks", which provide basic analytic capabilities, and a mechanism for combining these basic operations together in new ways to extend those capabilities. These two components are complementary. The choice of the basic analytic routines is influenced by how useful they will be in defining further analyses. Thus, a routine which computes a multiple regression and prints the result is not a useful building block, as nothing else can be done with it. One that removes variance components from a covariation matrix and returns the new matrix as a result would be very useful, as many common statistics could be defined with it. It is this two part approach, the interaction of the basic tools and the composition mechanism, which constitutes the essential difference between IDL and a conventional system."

I made use of the flexible, user-oriented aspects of IDL in defining the percentage agreement statistic used in chapter 2, and in combining basic statistical functions such as regression and analysis of variance for the various statistical tests in chapter 3. With the exception of minor adjustments in formatting, virtually all the statistics presented here are exactly as they were printed out by IDL.

The ease of experimentation which the interactive flexibility of IDL provides is a significant advance over more traditional statistics packages. Rather than forcing me into one of a limited number of modes of analysis, IDL allowed me to refine and combine methods to tailor the analysis to the particular details of my experiment. I hope the results in terms of increased relevance and readability are apparent.

C.3 The digital speech editor

This section describes the computer based interactive system for the display and reproduction of speech in digital form which I created and used to obtain the duration data discussed in chapter 3. As was the system described in C.1, it was written in Interlisp and makes use of video display facilities developed at Xerox PARC.

Figures 1 and 2 on the following pages are snapshots of the display screen showing what the speech looked like. The system displayed a digitized version of the time-amplitude waveform of the data segment, sampled at 7659 eight bit samples per second. It enabled me to place marks in the displayed speech signal, using the hand-held pointing device mentioned above in C.1. These marks show up as vertical lines on the display, with the time relative to the beginning of the data segment at the bottom, and the value of the signal, plus a comment I typed in, at the top. Thus the first mark in Figure 1 occurred 10.983 seconds into the text. At that point the speech signal was -1 in amplitude, and I judged that this was the beginning of the syllable *af*, as indicated by the notation. The scale of the display was variable, and Figure 1 is compressed 10 to 1, to give a broad overview of a section of the data. Figure 2 is uncompressed, and shows a short section of the data in detail.

The time-amplitude waveform is not as easy a basis for segmentation as, say, a formant trace would be, especially without acoustic feedback. Unfortunately the computational resources available to me were not sufficient to compute formants in a reasonable amount of time. To improve the accuracy of my segmentation, I therefore added to the system the ability to play back sections of the data between marks. By moving a mark back and forth and comparing the results by listening, I could quickly and easily determine the location of syllable boundaries. By the time I got to the end of the data segment, the system and my facility with the task had improved to the point that I needed between ten and fifteen minutes to segment a second of speech into syllables. Given the significant amount of silence on the tape, and the fact that both my performance and the system's facilities improved over time, it took me about 30 hours to segment the entire data segment, exclusive of the time which went into building the system.

The only significant difficulty I had with the transcription, as mentioned above at the beginning of chapter 3, was in discriminating unvoiced fricatives and sibilants from silence. This was the result of the relatively low sampling rate, which effectively filtered out all components of the signal above 3800 hertz, which is where most of the energy of these segments lies. Careful listening to the original text and extrapolation therefrom were sometimes needed to determine the correct placement of the boundary between, say, final [s] and a following pause, but there were few enough cases of this that even if some error was introduced as a result, it will not have affected the overall results significantly.

As I said in chapter 3, I think the end result of the segmentation using this combination visual and aural system was quite good, with few if any errors of more than one glottal wave, or 10 milliseconds.

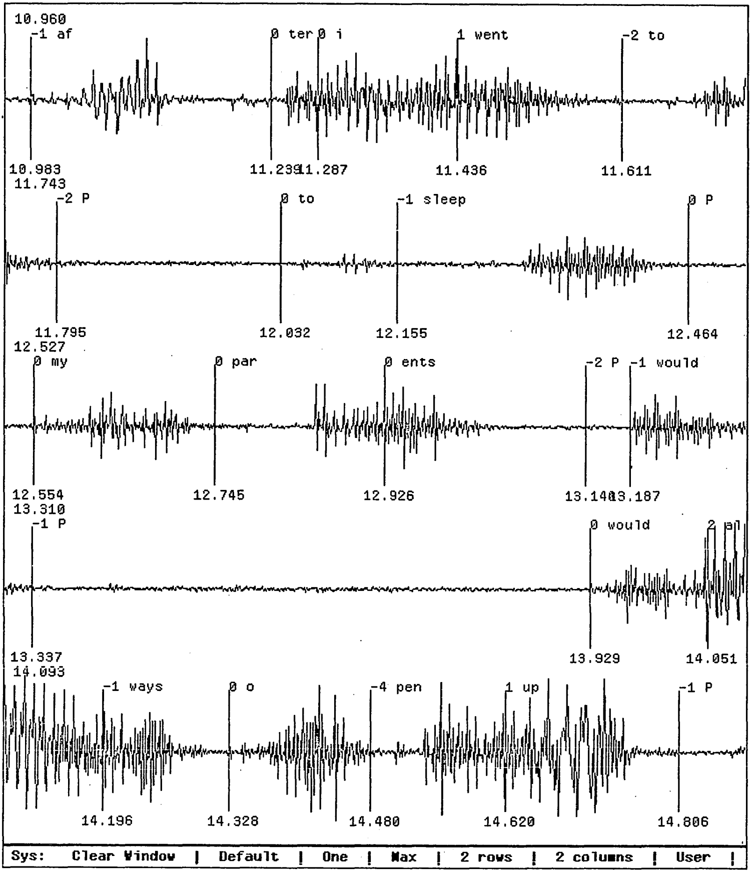


Figure 1. Snapshot of speech editing session with 4 seconds of the data segment displayed, compressed 10 to 1

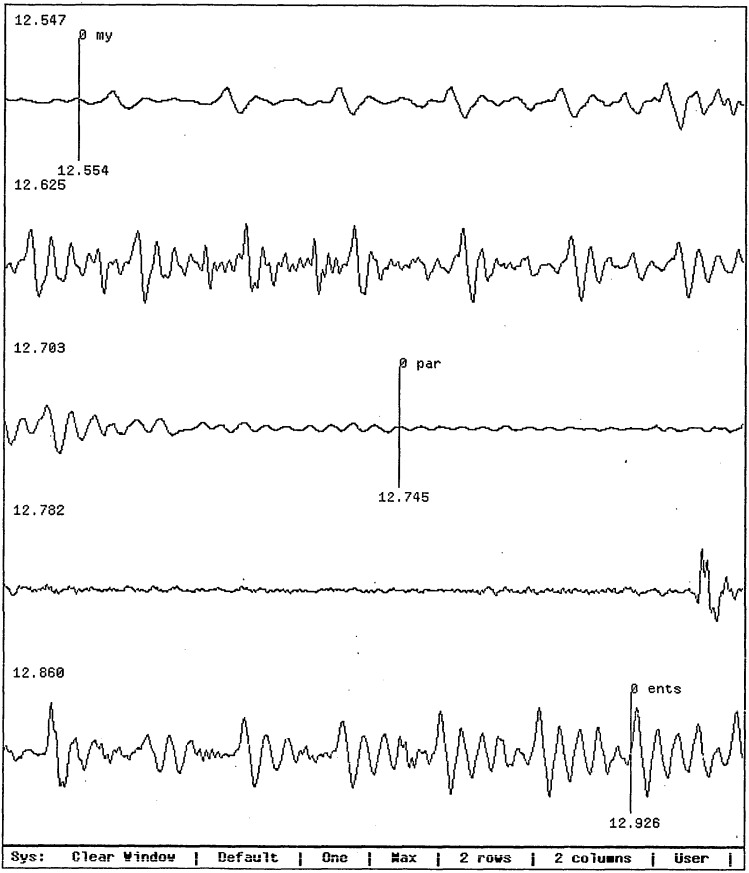

12.547
0 my
12.554
12.625
12.703
0 par
12.745
12.782
12.860
0 ents
12.926
Sys: Clear Window | Default | One | Max | 2 rows | 2 columns | User |

Figure 2. Snapshot of speech editing session with .5 seconds of the data segment displayed, uncompressed

References

What follows is by no means an exhaustive listing of relevant works. [Ladd 1978] has a good coverage of more recent work, but [Crystal 1969] remains the most complete bibliography available.

Abercrombie, David, D. B. Fry, P. A. D. McCarthy, N. C. Scott and J. L. M. Trim, (eds.) 1964. *In Honour of Daniel Jones*, London: Longmans.

Abercrombie, David 1964. "Syllable Quantity and Enclitics in English," in [Abercrombie et al. 1964].

Allen, Jonathan 1978. "The Role of Structural Constraints in Speech Production and Perception," *ms.*, University of Massachusetts Workshop on the Mental Representation of Phonology, Amherst MA.

Allen, James 1979. *A Plan-Based Approach to Speech Act Recognition*, Toronto: Technical Report 131/79, Dept. of Computer Science, Univ. of Toronto.

Anderson, J., J. Laver and T. Myers (eds.) to appear. *The Cognitive Representation of Speech*, Amsterdam: North Holland.

Baker, C. L. 1979. "Remarks on Complementizers, Filters, and Learnability," *ms.*, Sloan Foundation Workshop on Criteria of Adequacy for a Theory of Language, Stanford University.

Beach, W. A., S. E. Fox and S. Philosoph, (eds.) 1977. *Papers from the Thirteenth Regional Meeting*, Chicago: Chicago Linguistic Society.

Bell, Alan 1977. "Accent placement and perception of rhythmic structures," in [Hyman 1977].

Bolinger, Dwight 1958. "A Theory of Pitch Accent in English," *Word*, 14: 109-49, reprinted in [Bolinger 1965a].

——————— 1965a. *Forms of English: Accent, Morpheme, Order*, Cambridge MA: Harvard Univ. Press.

——————— 1965b. "Pitch Accent and Sentence Rhythm," in [Bolinger 1965a].

——————— 1972. "Accent is Predictable (if you're a Mind-Reader)," *Language*, 48: 633-44.

Bresnan, Joan 1978. "A Realistic Transformational Grammar," in [Halle et al. 1978].

Chafe, Wallace L. 1974. "Language and Consciousness," *Language*, 50: 111-133.

——————— 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Points of View," in [Li 1976].

——————— 1977. "The recall and verbalization of past experience," in [Cole 1977].

——————— 1979a. "The flow of thought and the flow of language," in [Givôn 1979].

——————— 1979b. "Some Reasons for Hesitating," in [Dechert and Raupach 1979].

Chase, W. G. (ed.) 1973. *Visual Information Processing*, New York, NY: Academic Press.

Chiarello, Christine, J. Kingston and E. E. Sweetser et al., (eds.) 1979. *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistic Society*, Berkeley CA: BLS.

Chomsky, Noam 1971. "Deep Structure, Surface Structure, and Semantic Interpretation," in [Steinberg and Jacobovits 1971].

Chomsky, Noam, and Morris Halle 1968. *The Sound Pattern of English*, New York: Harper and Row.

Clark, E. V., and H. H. Clark 1977. *Psychology and Language*, New York: Harcourt Brace Jovanovich.

Cogen, C., H. Thompson, G. Thurgood, K. Whistler and J. Wright, (eds.) 1975. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, Berkeley CA: BLS.

Cohen, P. R. 1978. *On knowing what to say: Planning speech acts*, Toronto: PhD dissertation, Dept. of Computer Science, Univ. of Toronto.

Cole, R. W. (ed.) 1977. *Current Issues in Linguistic Theory*, Bloomington IN: Indiana University Press.

Crystal, David 1969. *Prosodic Systems and Intonation in English*, Cambridge: Cambridge University Press.

Dechert, H. W., and M. Raupach (eds.) 1979. *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler*, The Hague: Mouton.

van Dijk, T. A. 1972. *Some Aspects of Text Grammars*, The Hague: Mouton.

El-Menoufy, A. M. E-S. 1969. *A Study of the Role of Intonation in the Grammar of English*, London: Doctoral dissertation, Dept. of General Linguistics, University College London.

Frazier, L., and J. D. Fodor 1978. "The sausage machine: a new two-stage parsing model," *Cognition*, 6: 291-325.

Fry, D. B. 1955. "Duration and Intensity as Physical Correlates of Linguistic Stress," *Journal of the Acoustical Society of America*, 27: 765-769.

——————— 1958. "Experiments in the Perception of Stress," *Language and Speech*, 1: 126-52.

Givón, Talmy 1976. "Topic, Pronoun, and Grammatical Agreement," in [Li 1976].

Givón, Talmy (ed.) 1979. *Discourse and Syntax*, New York: Academic Press.

Grimes, J. The Thread of Discourse. *1972*, Ithaca NY: Technical Report 1, Dept. of Modern Languages and Literatures, Cornell Univ..

Grossman, R. E., L. J. San and T. J. Vance (eds.) 1975. *Papers from the Parasession of Functionalism*, Chicago: Chicago Linguistic Society.

Gunter, Richard 1974. *Sentences in Dialog*, Columbia SC: Hornbeam Press.

Halle, Morris, Joan Bresnan and George A. Miller 1978. *Linguistic Theory and Psychological Reality*, Cambridge MA: MIT Press.

Halliday, M. A. K. 1967a. *Intonation and Grammar in British English*, The Hague: Mouton.

——————— 1967b. "Notes on Transitivity and Theme in English (Part II)," *Journal of Linguistics*, 3: 199-244.

Haviland, S. E., and H. H. Clark 1974. "What's New? Acquiring New Information as a Process in Comprehension," *Journal of Verbal Learning and Verbal Behavior*, 13: 512-521.

Hankamer, J., and I. A. Sag 1976. "Deep and Surface Anaphora," *Linguistic Inquiry*, 7: 391-426.

Householder, Fred 1957. "Accent, Juncture, Intonation, and my Grandfather's Reader," *Word*, 13: 234-45.

Hyman, Larry M. (ed.) 1977. *Studies in Stress and Accent*, Los Angeles: Southern California Occasional Papers in Linguistics No. 4, Dept. of Linguistics, Univ. of Southern California.

Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar*, Cambridge MA: MIT Press.

James, D. M. 1973. *The syntax and semantics of some English interjections*, Ann Arbor: Doctoral dissertation, Dept. of Linguistics, Univ. of Michigan.

Jespersen 1924. *The Philosophy of Grammar*, New York: W. W. Norton, 1965.

Kaplan, Ronald M. 1975. "On Process Models for Sentence Analysis," in [Norman and Rumelhart 1975].

Kaplan, R. M., B. A. Sheil and E. R. Smith 1978. *The Interactive Data-analysis Language Reference Manual*, Palo Alto CA: Xerox Palo Alto Research Center SSL-78-4.

Kintsch, W. 1974. *The Representation of Meaning in Memory*, Hillsdale NJ: Lawrence Erlbaum.

Labov, W., and D. Fanshell 1977. *Therapeutic Discourse*, New York: Academic Press.

Ladd, Dwight Robert, Jr. 1978. *The Structure of Intonational Meaning*, Ithaca NY: Doctoral dissertation, Linguistics Dept., Cornell Univ., available from University Microfilms International, Ann Arbor MI.

Lakoff, George 1970. "Global Rules," *Language*, 46: 627-639.

Lakoff, G., and Henry Thompson 1975a. "Introducing Cognitive Grammar," in [Cogen et al. 1975].

———————— 1975b. "Dative Questions in Cognitive Grammar," in [Grossman et al. 1975].

Lea, Wayne A. 1973. "Perceived stress as the 'standard' for judging acoustical correlates of stress," *ms.*, Acoustical Society of America, 86th Meeting.

——————— 1976. "Perceived stress patterns in selected English phrase structures," *ms.*, American Association of Phonetic Sciences, 1976 Annual Meeting.

——————— 1977. "Acoustic correlates of stress and juncture," in [Hyman 1977].

Legaly, M. W., R. A. Fox and A. Bruck, (eds.) 1974. *Papers from the Tenth Regional Meeting*, Chicago: Chicago Linguistic Society.

Lehiste, I. 1973. "Rhythmic Units and Syntactic Units in Production and Perception," *ms.*, Acoustical Society of America, 85th Meeting.

Li, Charles N. (ed.) 1976. *Subject and Topic*, New York: Academic Press.

Li, Hughes and Snow 1973. *Comparison of Stress Marking by Linguist and by Computer*, West Lafayette IN: School of Electrical Engineering, Purdue Univ.

Liberman, Mark 1975. *The Intonational System of English*, Cambridge MA: Doctoral dissertation, Dept. of Linguistics, MIT, available from IU Linguistics Club, Blomington, IN.

Liberman, Mark, and Alan Prince 1977. "On Stress and Linguistic Rhythm," *Linguistic Inquiry*, 8: 249-336.

Liberman, Mark, and Ivan Sag 1974. "Prosodic Form and Discourse Function," in [Legaly et al. 1974].

Liberman, M., and Henry Thompson 1980. "On Intonational Meaning: Some thoughts about washing the car," *ms.*, Stanford Sloan Workshop on Intonation.

Lieberman, Philip 1965. "On the Acoustic Basis of the Perception of Intonation by Linguists," *Word*, 21: 40-54.

Linde, Charlotte 1974. *The Linguistic Encoding of Spatial Infonnation*, New York: Doctoral dissertation, Columbia Univ.

Longacre, R. E. 1977. "Tagmemics as a Framework for Discourse Analysis," *Proceedings of the Second Annual Linguistic Metatheory Conference*, Dept. of Linguistics, Michagan State University.

Morgan, Jerry 1978. "Toward a Rational Model of Discourse Comprehension," *Theoretical Issues in Natural Language Processing - 2, Proceedings*, New York: ACM.

Newell, A. 1973. "Production Systems: Models of Control Structures," in [Chase 1973].

Nichols, Johanna 1979. "The meeting of East and West: confrontation and convergence in contemporary linguistics," in [Chiarello et al. 1979].

Norman, D. A., and D. E. Rumelhart 1975. *Explorations in Cognition*, San Francisco: W. H. Freeman.

O'Connor, J. D., and J. F. Arnold 1961. *Intonation of Colloquial English*, London: Longmans.

O'Malley, M. H., D. R. Kloker and D. Dara-Abrams 1973. "Recovering Parentheses from Spoken Algebraic Expressions," *IEEE Transactions on Audio and Electroacoustics*, AU-21: 217-220.

Petofi, J. 1973. "Towards an Empirically Motivated Grammatical Theory of Verbal Texts," in [Petofi and Rieser 1973].

Petofi, J. S., and H. Rieser (eds.) 1973. *Studies in Text Grammar*, Dordrecht: Reidel.

Perrault, C. R., and P. R. Cohen 1977. *Planning speech acts*, Toronto: AI-Memo 77-1, Dept. of Computer Science, Univ. of Toronto.

Pierrehumbert, J. 1979. "Intonation synthesis based on metrical grids," *ms.*, Acoustical Society of America, 97th Meeting.

——————— 1980. "A Two-tone, Autosegmental approach to English Intonation," *ms.*, Stanford Sloan Workshop on Intonation.

Pike, Kenneth L. 1945. *The Intonation of American English*, Ann Arbor: University of Michigan Press.

Postal, P. M., and G. K. Pullum 1978. "Traces and the Description of English Complementizer Contraction," *Linguistic Inquiry*, 9: 1-29.

Potter 1961. *Fundamentals of Music*, Mattapan MA: Gamut Music.

Propp, V. 1968. *Morphology of the Folktale*, English translation, 2nd edition, Austin: University of Texas Press.

Reinhart, T. 1976. *The Syntactic Domain of Anaphora*, Cambridge MA: Doctoral dissertation, Dept. of Linguistics, MIT.

Rubin, A. D. 1978. "A theoretical taxonomy of the differences beween oral and written language," in [Spiro et al. 1978], also available as Center for the Study of Reading Technical Report No. 35, Univ. of Illinois at Urbana-Champaign, 1977

Sag, I. A. 1976. *Deletion and Logical Form*, Cambridge MA: Doctoral dissertation, Dept. of Linguistics, MIT, available from IU Linguistics Club, Blomington, IN.

————— forthcoming. *The Syntax and Semantics of Verb Phrase Deletion*, New York: Elsevier North-Holland.

Sag, I. A., and S. Weisler 1979. *Temporal connectives and logical form*, Chiarello et al..

Schmerling, Susan F. 1976. *Aspects of English Sentence Stress*, Austin: University of Texas Press.

Selkirk, E. O. 1979. "On the Nature of Phonological Representation," in [Anderson et al. to appear].

————— forthcoming. *Phonology and Syntax: The Relation of Sound and Structure*, Cambridge MA: MIT Press.

Sgall, Petr, E. Hajičová and E. Benešová 1973. *Topic, Focus, and Generative Semantics*, Kronberg GDR: Scriptor Verlag.

Spiro, R., B. Bruce and W. Brewer, (eds.) 1978. *Theoretical issues in reading comprehension*, Hillsdale NJ: Lawrence Erlbaum.

Sproull, Robert F. 1979. *Raster Graphics for Interactive Programming Environments*, Palo Alto CA: Xerox Palo Alto Research Center CSL-1979-6.

Steinberg, D., and L. Jakobovits (eds.) 1971. *An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, Cambridge: Cambridge University Press.

Sternberg, S. 1969. "Memory-scanning: Mental processes revealed by reaction-time experiments," *American Scientist*, 4: 421-457.

Teitelman, W. 1978. *Interlisp Reference Manual*, Palo Alto CA: Xerox Palo Alto Research Center.

Thompson, Henry 1976. "Towards a Model of Language Production: Linguistic and Computational Foundations," *Statistical Methods in Linguistics*, 1976: 110-126.

——————— 1977. "Strategy and Tactics: A Model for Language Production," in [Beach et al. 1977].

——————— 1978. *Interim Interlisp Display Facility*, Palo Alto CA: Xerox Palo Alto Research Center internal memo.

Trager, George L., and Henry Lee Smith 1951. *Outline of English Structure*, Norman OK: Battenburg Press.

Vanderslice, Ralph, and Peter Ladefoged 1972. "Binary Suprasegmental Features and Transformational Word-Accentuation Rules," *Language*, 48: 819-38.

Wilensky, R. 1978. *Understanding Goal-Based Stories*, New Haven: Computer Science Research Report 140, Yale Univ.