in Intentions in Communication
ed. Cohen, Morgan & Pollack
MIT Press, 1990.

tanding computers and cognition: A new

mood in typologically differing lan-
enth International Congress of Linguists.

Chapter 14

The Meaning of Intonational Contours in the Interpretation of Discourse

Janet Pierrehumbert and Julia Hirschberg

1 *Introduction*

Recent investigations of the contribution that intonation makes to overall utterance and discourse interpretation promise new sources of information for the investigation of long-time concerns in natural-language processing. In Hirschberg and Pierrehumbert 1986 we proposed that intonational features such as *phrasing, accent placement, pitch range,* and *tune* represent important sources of information about the *attentional* and the *intentional* structures of discourse.[1] In this paper we examine the particular contribution of choice of tune, or *intonational contour,* to discourse interpretation. In particular, we propose that a speaker (S) chooses a particular tune to convey a particular relationship between an utterance, currently perceived beliefs of a hearer or hearers (H), and anticipated contributions of subsequent utterances. We claim that these relationships are compositional—composed from the *pitch accents, phrase accents,* and *boundary tones* that make up tunes. We further propose that the different aspects of tune meaning can be associated with different phonological domains. We assume the *intonational phrase* as our primary unit of meaning analysis.

In the following discussion we put forward a first approximation of a compositional theory of tune interpretation, together with the phonological assumptions on which it is based and the evidence from which we have drawn our proposals. We assume Pierrehumbert's (Pierrehumbert 1980; Beckman and Pierrehumbert 1986a) theory of intonational description, which we describe in sections 2–3. In section 4 we present our general approach to intonational meaning. In sections 5–7 we present the data upon which we base this account. In section 8 we explore avenues of further development for the theory and discuss implications for the study of discourse.

2 *Dimensions of Intonational Variation*

2.1 *Preliminaries*

In describing intonation patterns, we distinguish *stress, tune, phrasing,* and *pitch range. Stress* refers to the rhythmic pattern or relative prominence of

syllables in an utterance. *Tune* is the abstract source of fundamental frequency patterns—the difference between a typical declarative intonation and a question intonation is a tune difference. English has a very rich tune system, as the reader can appreciate by producing a monosyllable with many different intonation patterns. *Phrasing* refers to how a complex utterance is divided up. Each *intonational phrase* provides an opportunity for a new choice of tune, and as we will show, some parts of the tune serve to mark the *phrase boundaries*. Phrase boundaries are also indicated by the *duration* pattern and by pausing. *Pitch range* controls the graph paper on which the tunes are realized. One may increase one's pitch range for many reasons—for example, to project one's voice or to highlight the information in a particular phrase.

2.2 *Stress*

The *stress pattern* of an utterance is the pattern of relative prominence of the syllables. Word stress is assigned by lexical-phonological rules. Stress within the phrase is affected by considerations of information structure. For example, the following sentence would usually be produced with the main phrasal stress (the *nuclear stress*) on the word *vitamins*:

(1) Legumes are a good source of VITAMINS.

However, the nuclear stress would fall on *good* in a context where *sources of vitamins* are already under discussion, as in (2):

(2) A: Legumes are a pretty poor source of vitamins.
B: No. Legumes are a GOOD source of vitamins.

Stress manifests itself in the duration, amplitude, and spectral characteristics of the speech segments. In general, syllables with greater stress are more fully articulated than syllables with less stress. Stress pattern is independent of tune, in the sense that a given tune can be applied to materials with many different stress patterns and a given stress pattern can be produced with many different tunes. For example, (1) can be produced either with a falling-rising fundamental frequency (f_0) pattern on *vitamins* or with a rising pattern. These two possibilities are illustrated by the f_0 contours in figures 14.1 and 14.2.

Either pattern can also be applied to the same sentence when the nuclear stress is shifted to *good*, as in (2). Figures 14.3 and 14.4 show the two outcomes in this case.

2.3 *Tune and Phrasing*

In Pierrehumbert's system of intonational description, tunes are described as sequences of *low* (**L**) and *high* (**H**) tones, which determines the shape of the f_0 contour. Some of these tones (the ones participating in pitch accents)

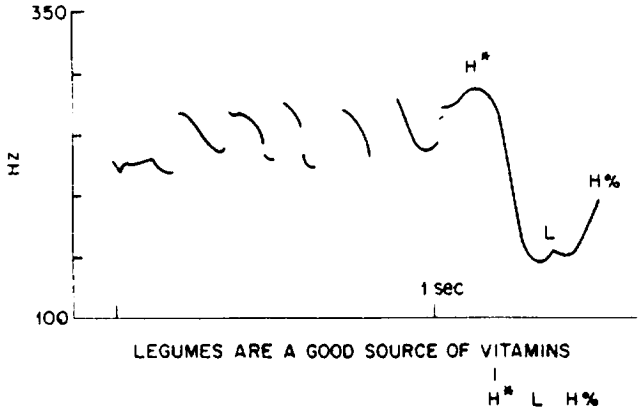


Figure 14.1
Falling-rising pattern on *vitamins*. Reprinted from Pierrehumbert 1980.

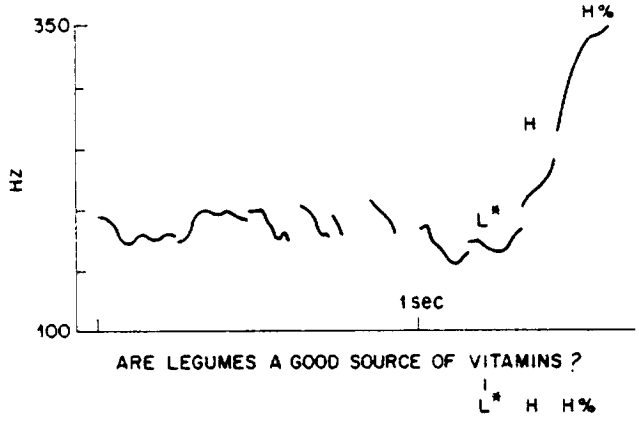


Figure 14.2
Rising pattern on *vitamins*. Reprinted from Pierrehumbert 1980.

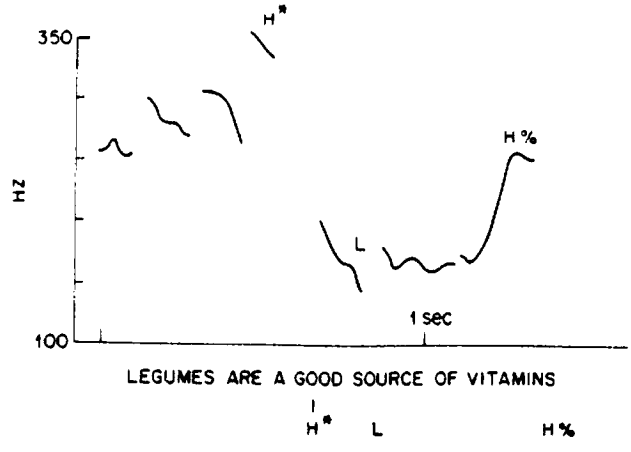


Figure 14.3
Nuclear stress on *good* with a falling-rising pattern. Reprinted from Pierrehumbert 1980.

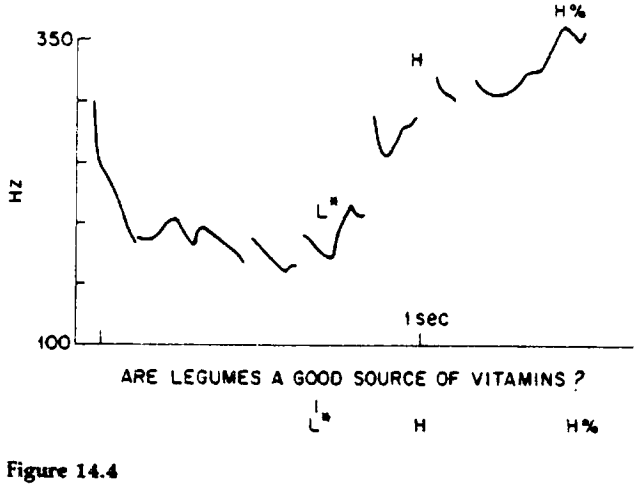
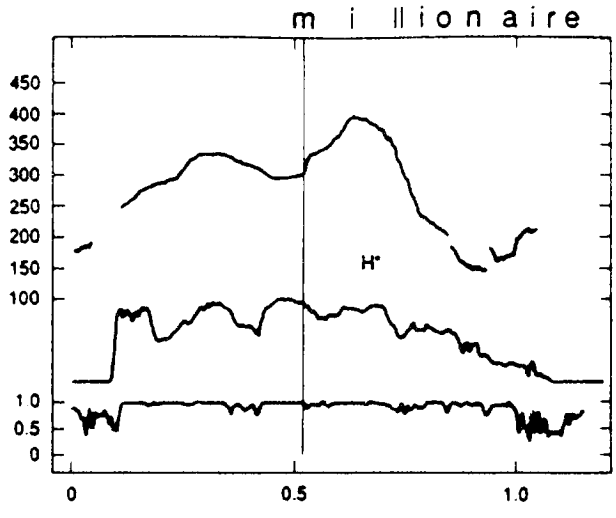


Figure 14.4
A rising pattern with nuclear stress on *good*. Reprinted from Pierrehumbert 1980.



humbert 1980.

Figure 14.5
A H* accent on *millionaire*. Reprinted from Pierrehumbert and Steele, in press.

go with stressed syllables. If the stress pattern for a given sentence is changed, the number and location of pitch accents is changed accordingly. Other tones, the *phrasal tones*, mark the edges of phonological phrases. If the way a sentence is divided into phrases is modified, the number and location of phrasal tones is changed.

Pitch accents mark the lexical item with which they are associated as prominent. There are six different types of pitch accent in English (Beckman and Pierrehumbert 1986a): two simple tones—high and low—and four complex ones. The high tone, the most frequently used accent, comes out as a peak on the accented syllable. It is represented as **H***. The "**H**" indicates a high tone, and the "*" that the tone is aligned with a stressed syllable. **L*** accents occur much lower in the pitch range than **H*** and are phonetically realized as local f_0 minima. The other English accents have two tones, of which one is selected to align with the stress. Using the diacritic "*" to indicate this alignment, these accents can be represented as **L*+H**, **L+H***, **H*+L**, and **H+L***. Accents with two like tones do not exist. Figures 14.5 and 14.6 illustrate the contrast between **H*** and **L+H***. The utterance in both cases is *Only a millionaire*, with the word stress for *millionaire* on the first syllable. The vertical line in the figure indicates the release of the [m] into the vowel. Note that both contours have an f_0 peak on the first syllable of *millionaire*. But there is a pronounced valley before the peak in the case of the **L+H*** accent. Figure 14.7 continues the comparison by illustrating **L*+H** on the same phrase. Now the low f_0

: 1980.

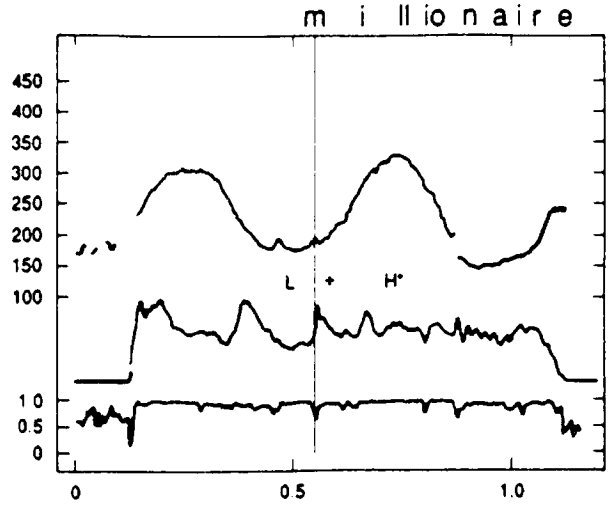


Figure 14.6
A L + H* accent on *millionaire*. Reprinted from Pierrehumbert and Steele, in press.

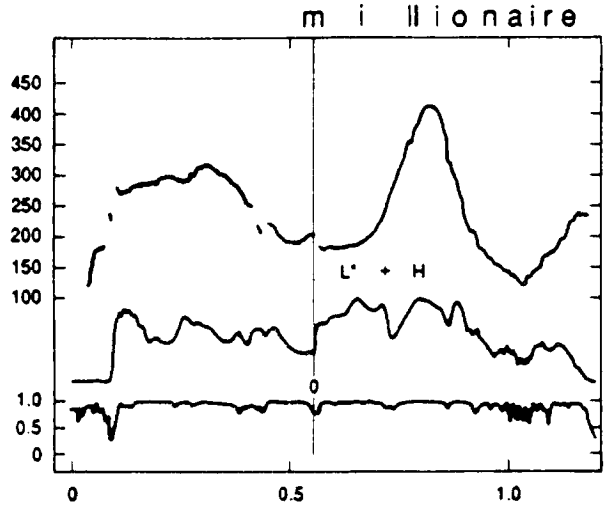


Figure 14.7
A L* + H accent on *millionaire*. Reprinted from Pierrehumbert and Steele, in press.

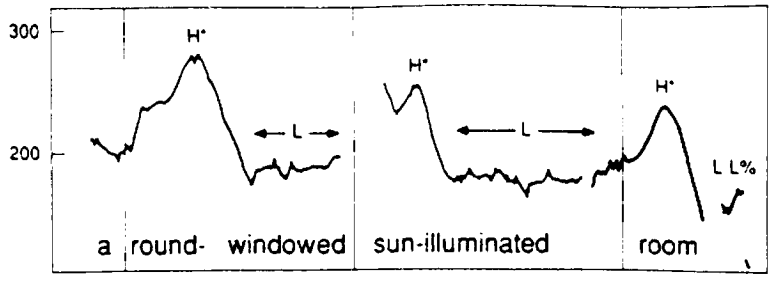


Figure 14.8
An intonational phrase with two intermediate phrases. Reprinted from Beckman and Pierrehumbert 1986a.

value continues past the [m]-release, and the peak occurs in the second syllable.

Beckman and Pierrehumbert 1986a report that two levels of phrasing in English are involved in the specification of tune. These are the *intermediate phrase* and the *intonational phrase*. A well-formed intermediate phrase consists of one or more pitch accents, plus a simple high or low tone (either **H** or **L**), which marks the end of the phrase. Continuing somewhat obsolete terminology from Pierrehumbert 1980, we will refer to this tone as the *phrase accent*. An important phonetic property of the phrase accent is that it controls the f_0 between the last pitch accent of the intermediate phrase and the beginning of the next intermediate phrase—or the end of the utterance. This is illustrated in figure 14.8, where the **L** phrase accent of each of the first two intermediate phrases shows its influence over an extended region. Vertical lines in the figure mark the phrase boundaries, as determined from phonetic segmentation of the utterance.

Intonational phrases are composed of one or more intermediate phrases. The end of an intonational phrase is marked with an additional **H** or **L** tone, which we will refer to as the *boundary tone* and indicate with the diacritic "**%**." This tone falls exactly at the phrase boundary. Since the end of every intonational phrase is also the end of an intermediate phrase, there are altogether four ways that the tune can go after the last pitch accent of an intonational phrase: **L L%**, **H L%**, **L H%**, and **H H%**.

A phrase's tune or *melody* is defined by its particular sequence of pitch accent(s), phrase accent(s), and boundary tone. Thus, an ordinary declarative pattern with a final fall is represented as **H* L L%**, a tune with a **H*** pitch accent, a **L** phrase accent, and a **L%** boundary tone. A typical interrogative contour is represented as **L* H H%**. (The contrast between these two melodies was illustrated in figures 14.1–14.4.)

Intermediate and intonational phrases can be identified by pausing and phrase-final syllable lengthening as well as by the extra melodic elements

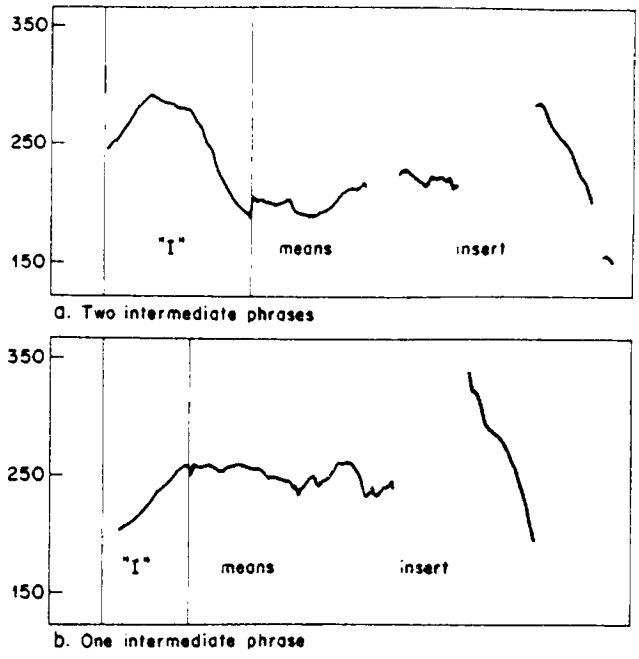


Figure 14.9
Alternate phrasings of *"I" means insert*. Reprinted from Beckman and Pierrehumbert 1986a.

of phrase accent and boundary tone present at the end. Figure 14.9 shows a sentence produced in two ways, once with an intermediate phrase boundary after *I* and once as a single intermediate phrase. Note that *I* carries an f_0 fall in (a) and its duration (indicated by the vertical line) is greater than in (b).

2.4 Pitch Range

When S's voice is raised, the overall *pitch range*—the distance between the highest point in the f_0 contour and the *baseline* (the lowest point S realizes over all utterances)—is expanded. Thus, the highest points in the contour become higher and other aspects are affected proportionally. Figure 14.10 shows a simple utterance (the word *Anne*) produced in seven different overall pitch ranges with a **H* L L%** tune. The contours are similar in shape but differ in overall scaling, especially in the peak f_0 value.

In addition to variations in overall pitch range, the intonation system exploits a local time-dependent type of pitch range variation called *final lowering*. In the experiments reported in Liberman and Pierrehumbert 1984 it was found that the pitch range in declaratives is lowered and compressed in anticipation of the end of the utterance. Final lowering begins about

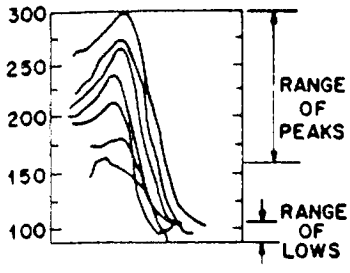


Figure 14.10
Anne produced with seven different pitch ranges. Reprinted from Liberman and Pierrehumbert 1984.

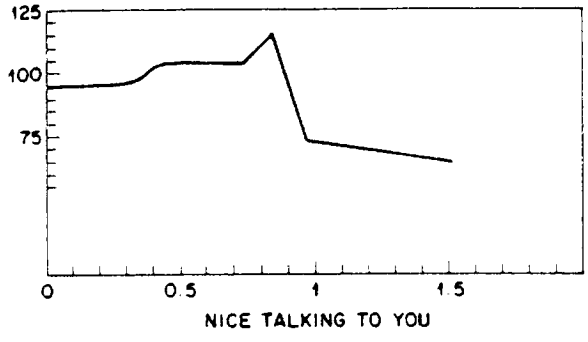


Figure 14.11
A phrase synthesized with final lowering. Reprinted from Hirschberg and Pierrehumbert 1986.

half a second before the end and gradually increases, reaching its greatest strength right at the end of the utterance.

Both overall pitch range and final lowering enter into intonational interpretation. They are especially important in conveying the hierarchical segmentation of the discourse. Many researchers have observed that the pitch range is expanded at the beginning of a new topic (Schegloff 1979; Brazil, Coulthard, and Johns 1980; Butterworth 1975). In Hirschberg and Pierrehumbert 1986 and Silverman 1987 it was also observed that final lowering reflects the degree of "finality" of an utterance; the more final lowering, the more the sense that an utterance "completes" a topic. Figures 14.11 and 14.12 illustrate this point with contours synthesized by the intonation synthesis program described in Anderson, Pierrehumbert, and Liberman 1984. The first sounds like the usual pronunciation of the sentence *Nice talking to you*, whereas the second creates a sense that S has reservations that are unexpressed for the sake of politeness. Recent experiments

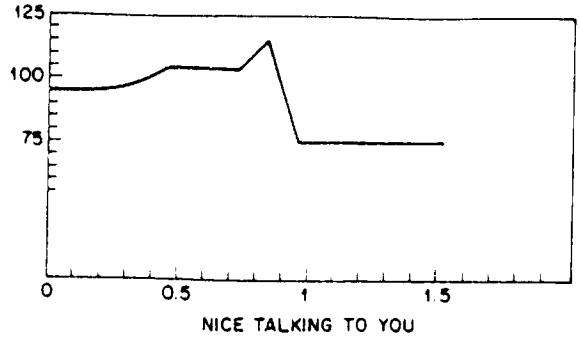


Figure 14.12
A phrase synthesized without final lowering. Reprinted from Hirschberg and Pierrehumbert 1986.

(Silverman 1987) show that pitch range and final lowering can function perceptually to disambiguate texts whose hierarchical structure is unclear.

In addition to its role in signaling overall discourse structure, pitch range interacts with the basic meanings of tunes to give their interpretation in context. For example, if S speaks up, S is likely to sound more assertive. Perceived assertiveness may make some derived interpretations of the tune seem more plausible than others.

3 *Tonal Realization*

The way that elements of the tune are mapped into f_0 values is discussed in detail in Pierrehumbert 1980 and Liberman and Pierrehumbert 1984. We mention two main effects here in order to help the reader interpret the examples that follow.

Upstep raises the boundary tone after a **H** phrase accent. The sequence **H H%** comes out as a high plateau followed by an additional rise at the very end. The sequence **H L%** comes out as a high plateau without any drop at the end. *Catathesis*, or *downstep*, lowers and compresses the pitch range after any of the two-tone accents. The rule applies iteratively, so that a succession of such accents creates a decending staircase in the f_0 pattern. It is important to note that catathesis affects a **H** phrase accent when one of the two tone accents occurs in nuclear position. The result is a kind of "mid" tone, lower than the preceding **H** tone but still well above the bottom of S's range. The effects of catathesis disappear at an intermediate phrase boundary; for each new intermediate phrase, a fresh selection of overall pitch range is made.

Figures 14.13–14.15 schematize the f_0 contours resulting for different combinations of pitch accent, phrase accent, and boundary tone. The first

H*

L*

L*+H

Figure 14.1
Schematic f_0

L + H*

H* + L

H + L*

Figure 14.14
Schematic f_0

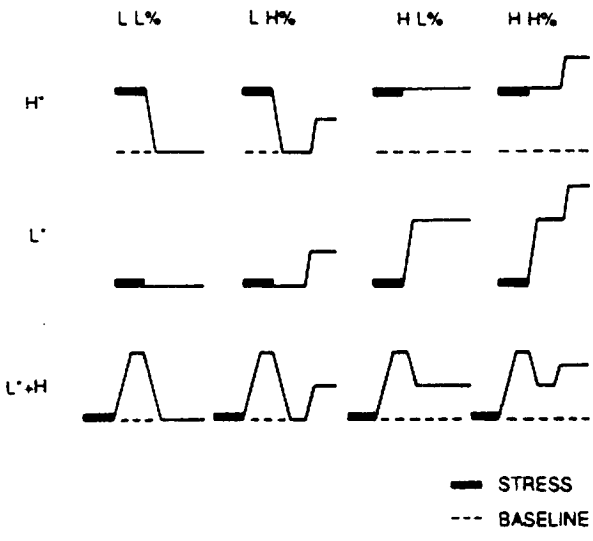


Figure 14.13
Schematic f_0 contours.

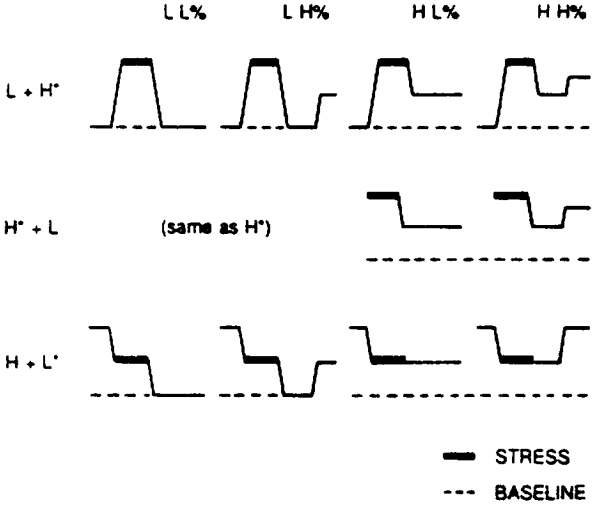


Figure 14.14
Schematic f_0 contours.

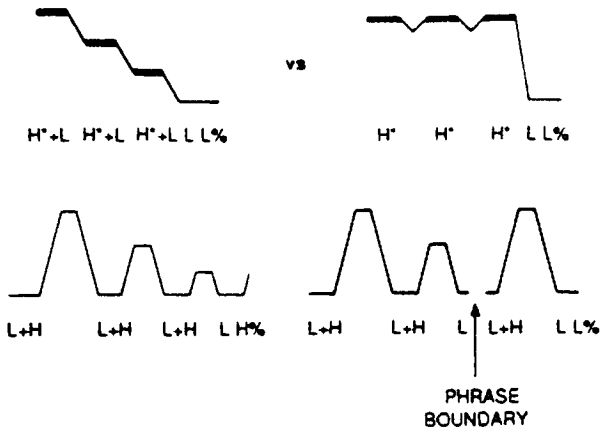


Figure 14.15
Schematic f_0 contours.

two figures provide an inventory of phrase-final configurations, and the third illustrates the operation of catathesis in sentences with several pitch accents.

3.1 Transcription and Theories of Intonational Meaning

The transcription system described here was originally motivated by phonetic and phonological considerations. It aimed at being comprehensive (by affording an analysis for all naturally occurring patterns) while at the same time exhibiting simplicity and symmetry in the abstract representation and the realization rules proposed. Of course, this is not enough. In the segmental domain, linguistic categories are expected to relate both to differences in sounds and articulations and to differences in semantic interpretation. For example, we say that [p] is different from [b] because they are pronounced differently, and because [pit] means something different than [bit] does.

Thus, any theory of transcription must be viewed as provisional unless it is supported by considerations both of sound structure and of interpretation. The transcription theory advances work on interpretation by suggesting what cases count as instances of the same category. Theories differ in the category structure they suggest. An incorrect theory can make it difficult to establish interpretations, by grouping together contours that actually hve disparate meanings or by drawing distinctions that have no meaning.

In view of this situation, we would like to call attention to some ways in which Pierrehumbert's transcription system differs from others in the way contours are cross-classified. First, Pierrehumbert proposes that the inven-

tory of pitch accents is the same in nuclear as in prenuclear position. Nuclear configurations differ from prenuclear ones because of the phrasal tones following the accent. This position contrasts with work in the British school (see, for example, Crystal 1969, O'Connor and Arnold 1961) in which the nuclear configuration is not decomposed, and a fundamental distinction is drawn between the nuclear and prenuclear inventories. If we are persuasive in our account of how pitch accents show stable meanings in different tonal contexts, then the British school work misses important generalizations.

Second, the primitives in the theory are tone levels rather than tone rises or falls. This permits us to describe **H* H H%** and **H* L L%** as involving the same pitch accent. In other approaches, such as those found in Bolinger 1958 and Gussenhoven 1983, these contours are entirely different because one is rising and the other is falling. A very strict dynamic tone theory is unable to differentiate among any of the rising contours (**H* H H%**, **H* H L%**, **L* H H%**, **L* H L%**, **L* L H%**).

Third, the theory has two tones rather than the four proposed in Pike 1945 and Liberman 1975. The reduction to two tones is made possible by using a catathesis rule to describe cases in which the f_0 contour shows a descending staircase of values. Transcription systems lacking a catathesis rule will in general draw excessive distinctions, from the point of view of interpretation.

The particular way the catathesis rule is formulated leads to some important partial similarities. The **H* + L H L%** pattern (which ends with a sustained "mid level tone") is analyzed with the same accent as the "stepping" declarative pattern **H* + L H* + L H* + L L L%**. We believe a common meaning can be identified across these two cases, as we will argue below. In other theories, this common meaning is not expected. In particular, in Ladd 1983 catathesis is not triggered by the left-hand context of a tone but is rather an independent feature. Using "!" for this feature, Ladd would give the transcriptions **H* !H** and **H* !H* !H* L**. In these transcriptions the two nuclear pitch accents are different. In Ladd's theory also, the contrast between **H* H L%** and **H* + L H L%** comes out as **H* H** versus **H* !H**. That is, he would be led to look for a difference in the interpretation of the phrase accent, where we would be looking for a pitch accent difference.

4 *The Interpretation of Tunes*

Past characterizations of the meanings of particular tunes have variously portrayed tune as conveying speaker attitude (O'Connor and Arnold 1961; Liberman 1975), such as politeness, deference, judiciousness, surprise, or seductiveness; emotion, such as hate or anger; speech acts (Sag and Liberman

1975; Liberman and Sag 1974), such as statements, requests, or contradictions: propositional attitudes (Ward and Hirschberg 1985), such as belief, ignorance, or uncertainty; presupposition and focus of attention (Jackendoff 1972; Ladd 1980); as well as less easily characterized meanings such as "up-in-the-airness" and "more to come." However, few of these characterizations have been successful for particular tunes, and none seems appropriate as a general approach to tune meaning.

Though speaker attitude may sometimes be inferred from choice of a particular tune, the many-to-one mapping between attitudes and tune suggests that attitude is better understood as derived from tune meaning interpreted in context than as representing that meaning itself. For example, as Ward and Hirschberg (1985) have shown, speaker uncertainty, incredulity, politeness, and irony can all be derived from the use of the **L*+H L H%** contour in different contexts. Similar problems, as well as experiments that have found pitch range and voice quality to be associated with perceived speaker emotions (Ladd et al. 1985), indicate that emotion is *not* a useful way of characterizing tune. Neither speech acts nor propositional attitudes—at least as standardly understood—appear to provide sufficient characterizations for available tunes in English. For example, the **H* L L%** tune used with simple declaratives is also frequently used with *wh*-questions. It is difficult to see how either a propositional attitude approach or a speech act analysis could produce a meaning for this contour that would accommodate both these common uses. In general, it seems advisable to divorce intonational meaning from speaker beliefs. For example, the **L*+H L H%** contour can be used to convey either that S believes **P** (as in (3)) or that S does not believe **P** (as in (4)):

(3) A: Who ordered the veal?
B: I'm having beef
L*+H L H%

(4) A: Here's your roast beef, sir.
B: I'm having beef
L*+H L H%
But I'm a vegetarian. There must be some mistake.

A more fruitful approach has been suggested by Gussenhoven (1983). He attempts to characterize the meaning of "nuclear tones" in terms of the status of information with respect to a shared "background," which is developed by speaker and hearer during the course of a conversation. We agree that this sort of information is part of what tunes convey. However, we disagree with the substance of Gussenhoven's description. His interpretations of particular tunes do not appear correct for American English. Also, the transcription system Gussenhoven proposes does not support

some of the generalizations we have noted and present below. Empirical counterevidence to some of Gussenhoven's claims appears in Ward and Hirschberg 1985 and Pierrehumbert and Steele 1987. Another account that treats intonational meaning in terms of the information status of accented items with respect to the discourse appears in Brazil, Coulthard, and Johns 1980.

Most analyses of the meaning of intonational contours to date have at least implicitly taken a holistic approach to tune meaning, confining the domain of interpretation to the phrase or utterance (see, for example, Bolinger 1958; O'Connor and Arnold 1961; Ladd 1980). This approach has probably been more a practical matter than a theoretical conviction: it is difficult to identify the "meaning" of parts of a contour until one has some idea of what the "meaning" of the whole might be. However, some of the individual tunes that have been successfully studied suggest that tune meaning is more usefully viewed as compositional. Tunes that share certain tonal features seem intuitively to share some aspects of meaning. For example, tunes such as **L*+H L H%**, **H* L H%**, and **L+H* L H%** that share a **L** phrase accent and a **H** boundary tone share also a sense that the current utterance will be completed by a subsequent utterance (Hirschberg and Pierrehumbert 1986). And various types of question contour, **L* H H%** and **H* H H%**, do share common high phrase accents and boundary tones while differing in the pitch accents used with them. A noncompositional approach fails to capture such generalities. However, the most ambitious attempt to provide a compositional account of intonational meaning (Pike 1945) was hindered by an inadequate representational system.

4.1 A Compositional Approach to Tune Meaning

We propose that speakers use tune to specify a particular relationship between the "propositional content" realized in the intonational phrase over which the tune is employed and the mutual beliefs of participants in the current discourse. Although the interpretation of any token of a tune type may vary along many other dimensions—voice quality, pitch range, as well as nonintonational features—any instance of a given tune will convey the same relationship. So, for example, any **H* L L%** tune will have in common with others the conveying of a certain relationship between the proposition realized by the phrase and propositions mutually believed in the discourse—whether that **H* L L%** tune is used with a *wh*-question or a syntactic declarative.

Following Clark and Marshall 1981 and Joshi 1982, we understand the *mutual beliefs* of a discourse to be those beliefs that conversational participants come to believe to be shared among them as a direct result of the conversational interaction. In particular, we make use of the notion of *one-sided* mutual belief—A's beliefs about what is mutually believed by A and

B. We will assume that a basic goal of a speaker S is to modify what (S believes) a hearer H believes to be mutually believed. For expository purposes, we will describe S's use of tune in terms of the intention to add to what (S believes) H believes to be mutually believed—or not—or to call attention to certain relationships between propositions realized by an utterance and other propositions that (S believes) H believes to be mutually believed. In this paper we will not specify how these intentions are related to the *intentional structure* of a discourse (Grosz and Sidner 1986). However, it seems clear that aspects of the intentional structure as well as the *attentional structure* of a discourse can be conveyed by choice of tune. For example, S may seek to inform H of some proposition x by communicating that x is to be added to what H believes to be mutually believed between S and H—via the tune S chooses. And S may seek to convey the information status of some item y—say, that y is old information that is to be treated as particularly salient—by the type of accent S uses in realizing y. Note in particular that S's beliefs are *not* specified by choice of tune—the "declarative" contour **H* L L%**, for example, will *not* be translated *S believes x*. But S's belief in x may be inferred from the combined meanings of pitch accents, phrase accents, and boundary tone, as they are used in particular contexts.

Our idea of the compositionality of tune meaning is based upon a hierarchical model of phonological domain, in which the scope of interpretation of tones is the node to which they are attached. So, the components of tune—pitch accents, phrase accents, and boundary tones—are each interpreted with respect to their distinct phonological domains.[2] Pitch accents, phrase accents, and boundary tones each operate on a (progressively higher) domain of interpretation. Not only is each of these types of tone interpreted over a distinct domain, but each contributes a distinct type of information to the overall interpretation of a tune.

Pitch accents convey information about the status of the individual discourse referents, modifiers, predicates, and relationships specified by the lexical items with which the accents are associated. For example, in (5) each **H*** provides information about predicates and arguments that are each denoted by a single lexical item—*train*, *leaves*, and *seven*—and how S intends these to be interpreted with respect to H's beliefs about their mutual beliefs:

(5) The train leaves at seven
H* H* H* L L%

Accenting or deaccenting of items in general appears associated with S's desire to indicate the relative salience of accented items in the discourse. The type of accent chosen conveys other sorts of information status. For example, accent type can indicate whether accented items or things predicated of them are to be included among items H believes mutually believed

or whether they should be excluded, whether something predicated of these items should be inferable from beliefs H already holds, or whether relationships in which S believes the items participate should be identified by H.

Phrase accents convey information at the level of the intermediate phrase. In (5) there is but a single intermediate phrase, with a **L** phrase accent. In (6), however, there are two:

(6) The train leaves at seven or nine twenty-five
H* H* H* H H* H* L L%

Here, the **H** phrase accent after *seven* has scope over the phrase *the train leaves at seven*, and the **L** phrase accent after *five* has scope over *or nine twenty-five*. We propose that S chooses phrase accent type to convey the degree of relatedness of one such phrase to preceding and succeeding intermediate phrases. Where a phrase like *the train leaves at seven* has a **H** phrase accent, for example, it is more likely to be interpreted as a unit with a phrase that follows.

The boundary tones contribute information about the intonational phrase as a whole. Whereas the domain of phrase accent and boundary tone is the same in (5), they differ in (6), where the **L%** contributes to the interpretation of the whole phrase *the train leaves at seven or nine twenty-five*. And whereas both (5) and (6) consist of single intonational phrases, the exchange in (7) has two:

(7) a. The train leaves at seven
H* H* H* L H%
b. It'll be on track four
H* H* L L%

We believe that boundary tones convey information about relationships among intonational phrases—in particular, about whether the current phrase is to be interpreted with particular respect to a succeeding phrase or not. This directionality may be further refined. It seems possible, for example, that the hierarchical and satisfaction-precedence reltionships that Grosz and Sidner (1986) propose as the bases of their intentional structure may be signaled by particular boundary tones.[3] So, in (7) S can indicate by a **H** boundary tone in (7a) that (7a) is to be interpreted with particular respect to a succeeding phrase (7b). In Grosz and Sidner's terms, it seems plausible to postulate a dominance relationship existing between (7a) and (7b)—the satisfaction of the purpose S has in uttering (7b) contributes to the satisfaction of S's purpose in uttering (7a) by further elaboration. So, in (7) the "forward reference" signaled by the boundary tone might be interpreted as indicative of a hierarchical relationship. Certainly with a **L** boundary tone in (7a), the relationship is less clearly marked. Consider the more

ambiguous exchange in (8):

(8) a. The train leaves at seven
H* H* H* L H%

b. There's a full moon tonight
H* H* H* L L%

With a H boundary tone in (8a), H will be much more likely to try to infer some relatinship between the state of the moon and the departure time of the train than if a L boundary tone is substituted.

So, we propose that tune meaning is composed of the meanings of three types of tone—pitch accents, phrase accents, and boundary tones—which have scope over three different domains of interpretation. Together, these intonational features can convey how S intends that H interpret an intonational phrase with respect to (1) what H already believes to be mutually believed and (2) what S intends to make mutually believed as a result of subsequent utterances. We believe that other characterizations of tune interpretation can in many cases be described in terms of the more general meanings we propose. For example, the conveyance of speaker attitudes like uncertainty or politeness or surprise, the conveyance of performatives like contradiction or declaration, and even turn-taking phenomena can be explained in terms of S's conveyance of various types of information status and propositional relationships—especially when combined with meanings conveyed by other intonational and nonintonational features.

The major support for our compositional approach to intonational meaning comes from an examination of how the different pitch accents are interpreted. In the following section we concentrate on examples in which the same pitch accent is used throughout the phrase and consider the contributions of accent, phrase accent, and boundary tone to the intonational meaning. Any success in identifying the meaning of different accents across different choices of phrase accent and boundary tone tends to support the idea that intonational meanings are compositional. Similarly, success in identifying the meaning of phrase accents when pitch accents and boundary tones are varied or identifying the meaning of boundary tones over the same comparisons also supports a compositional approach. Any success in deriving the varied meanings heretofore associated with particular melodies in different contexts tends to support our ideas about what the basic meanings of melodies can be like.

5 *The Interpretation of Pitch Accents*

All pitch accents render salient the material with which they are associated.[4] This is true regardless of the type of accent in question. In the phonological descriptions given in Liberman 1975 and Pierrehumbert 1980

this generalization arises because accents may be associated only with the most stressed material in the phrase: in Selkirk 1984 accents are taken to be prior, and accented material becomes stressed. Either way, salience goes with accent location and not with accent type. Accented material is salient not only phonologically but also from an informational standpoint. And items that are deaccented, by extension, do not undergo this salience marking—although they may already be salient or become salient by other means.

For purposes of illustration, we will view the logical form corresponding to an intonational phrase as an open expression in which accented items are replaced by variables.[5] What accentuation means, operationally, in this schema is that each variable has associated with it some indication of S's communication of the variable's information status with respect to what H believes to be mutually believed. This variable may or may not be instantiated with a representation of the accented item, depending upon the accent type employed. By this method, the utterance of (9) might be represented as shown in (10):

(9) George likes pie
H* **H* L L%**

(10) *x* likes *y*

x (**H***)
y (**H***)

x = George
y = pie

The open expression is *x likes y*. The instantiation of *x* is *George*, a pointer to an individual. The instantiation of *y* is *pie*, a pointer to a class. Both the individual and the class in question are marked as salient by the mere fact that the lexical items pointing to them are uttered with an accent. The further elaboration of the information status of both *George* and *pie* is indicated by the accent type employed—here, **H***—as we discuss below. In general, we believe that all accent types can be used to convey information to H about how the propositional content of the (perhaps partially) instantiated expression corresponding to the utterance is to be used to modify what H believes to be mutually believed.

5.1 *The H* Accent*

The **H*** accents above and in utterances in general convey that the items made salient by the **H*** are to be treated as "new" in the discourse. More generally, intonational phrases whose accents are all **H*** appear to signal to H that the open expression is to be instantiated by the accented items and

the instantiated proposition realized by the phrase is to be added to H's mutual belief space. When combined with a **L** phrase accent and a **L** or a **H** boundary tone, this is the pitch accent of "neutral declarative intonation." That is, it is appropriate when S's goal is to convey information. This contour may also be employed when S believes that H is already aware of the information, if S wishes to convey that it is mutually believed. For example:

(11) You turkey
H* L L%
You deliberately deleted my files
H* H* H* L L%

The **H*** accent can also combine with a **H** phrase accent and either a **H** or a **L** boundary tone. The first yields the so-called *high-rise question*. This pattern (**H* H H%**) may be used in preference to the standard *yes-no question contour* (**L* H H%**) when the questioned phrase simultaneously conveys information. Pierrehumbert (1980) notes the following examples. In the first example, due to Mark Liberman, he approaches a receptionist with a view to finding out if he is in the right place for his appointment and says,

(12) My name is Mark Liberman
H* H* H H%

In this case it seems that the entire phrase is intended to convey 'My name is Mark Liberman, and are you expecting me, or, am I in the right place?' That is, the **H*** accents convey that information is to be added to H's mutual beliefs, and the **H** phrase accent and boundary tone "question" the relevance of that information. In the second (naturally occurring) example a young woman was asked after a movie whether she liked the picture and replied,

(13) I thought it was good
H* H* H H%

This utterance might be glossed 'I thought it was good, but do you agree with me?' Again, S is providing information while asking for a comment on its appropriateness. In either of these cases it seems that a **L* H H%** contour would be infelicitous—and would probably convey that S was suffering from amnesia!

On syntactic yes-no questions, the contrast between the use of **H*** and the use of **L*** is somewhat less striking. However, **H* H H%** seems more often used when S believes that the answer to a question is yes—a confirmation question. For example, the authors of this paper were hard at work on it, when one wished to confer with the other and uttered,

(14) May I interrupt you
H* H H%

With this indirect speech act, it was clear that S thought it likely that an interruption would be permitted. **L* H H%**, on the other hand, conveys more of a sense that S is sincerely in doubt about the nature of the response. This is consistent with the view that S uses **H*** to try to add information to H's mutual belief space.

The comparison of **H* L L%** and **H* H H%** contours provides support for viewing intonational meaning in terms of attempted modifications of H's mutual beliefs. In both cases S attempts to establish that some particular information is shared. So, approaching tune meaning in terms of H's mutual beliefs permits a generalization of the **H*** meaning across both declarative and interrogative contexts.

The last case of the **H*** accent is in the *plateau contour*, currently being investigated by Hirschberg and Ward. This contour, **H* H L%**, has a peak on the accent syllable and then continues at the same high level. It is used to elaborate upon some previous statement—as to provide support or detail—as in (15):

(15) Wally: Mostly they just sat around and knocked stuff. You know.
The school
H* H L%
Other people
H* H L%

Here again instantiated expressions are to be added to H's mutual beliefs, although phrase accent and boundary tone indicate that the relationship of these expressions to other expressions realized in the discourse will differ from **H* L L%** and **H* H H%**.

The comparison of **H* L L%** and **H* H L%** with **H* H H%** provides support for viewing intonational meaning in terms of attempted modifications of H's mutual beliefs. In all cases S attempts to establish that some particular information is shared—by supplying that information for H or by attempting to elicit it. So, approaching tune meaning in terms of H's mutual beliefs permits a generalization of the **H*** meaning across both declarative and interrogative contexts.

5.2 *The L* Accent*

The **L*** accent marks items that S intends to be salient but not to form part of what S is predicating in the utterance. Schematically, one might say that S conveys that these items are not to be instantiated in the open expression that is to be added to H's mutual beliefs.

L* accents commonly appear in canonical yes-no questions—**L* H H%**—as we noted above. In questions like (16), for example, both *prunes* and *feet* are marked as salient by their **L*** accents:

(16) Do prunes have feet
L* **L* H H%**

However, S predicates nothing of these entities. In fact, S's motivation for marking these items as salient is the desire that H make such a predication. So, one common interpretation of the exclusion of salient items from the predication of an utterance is that S is not able to include them in some predication.

The **L* H H%** contour may also be used to convey incredulity. In such cases the **L*** accent's "salience-without-predication" may be interpreted as signaling that S believes the current instantiation of the open expression to be incorrect. An old Russian émigré joke relies on this usage. A staunch old Bolshevik is forced to confess publicly and reads as follows:

(17) I was wrong
L* **H** **L* H H%**
And Stalin was right
L* **H** **L* H H%**
I should apologize
L* **H** **L* H H%**

S may also employ **L*** accents when the instantiated expression is believed already part of H's mutual beliefs. For example, if S is asked to supply a list of things he wants for his birthday, when his desire for a Pavoni espresso machine is already mutually believed, he may begin,

(18) Well, I'd like a Pavoni ...
L* **L*** **L* L H%**

In this way, S conveys that his desire for this gift is already mutually believed by H. Such utterances may be made for the sake of completeness in listing, as a reminder, or to reassure H that he still wants a present that she has already purchased.

Of course, S may employ **L*** accents to convey this sense of existing mutual belief even when in fact S actually does not believe that this mutual belief exists. Such situations arise when S instructs, reprimands, or contradicts H, conveying that information should already be mutually believed even if it is not. For example, in (19) the use of the **L* L H%** pattern has a rather insulting effect, by suggesting that H should have had in mind something that she clearly did not.

(19) A: Let's order the Chateaubriand for two.
B: I don't eat beef
L* L* L H%

Exchanges like this led Liberman and Sag (1974) to label the L* L H% melody *contradiction contour*. However, this description is both too narrow and too broad. As Carlson (personal communication) points out, the melody cannot be used for just any sort of contradiction. It is only appropriate when S intends to convey that H should already be aware of what S is saying. It is not appropriate, for instance, in (20):

(20) A: My chances? The election isn't over till the last ballot has been counted.
B: #But CBS has just declared you the next president
L* L* L* L* L* L H%

In addition, the melody is used in many cases (such as (18)) where S is not in any sense contradicting H.

Additional evidence for our account of the meaning of the L* accent comes from the common use of this accent with lexical items that have, for independent reasons, been treated as extrapropositional, such as greetings, vocatives, and so-called cue phrases. For example, greetings such as (21) are commonly produced with L* accents:

(21) Good morning
L* L* L H%

In such cases it would be a mistake to account for the L* accent as associated with the conventionality of the statement. Conventional statements that are actually intended to convey information would not be likely to have L* accents. Consider the implausibility of continued employment for a switchboard operator who answered callers with (22):

(22) You have reached AT&T Bell Laboratories
L* L* L* L* L H%

Both preposed and postposed vocatives are frequently produced with L*, especially if S already has H's attention (see Beckman and Pierrehumbert 1986a for a discussion of the phonological analysis of these cases). Consider (23) and (24):

(23) Anna your lunch is ready
L* H H* H* L H%

(24) Your lunch is ready Anna
H* H* L L* L H%

A **H*** accent is possible on the preposed vocative if S does not already have H's attention, as in (25):

(25) Anna your lunch is ready
H* **L** **H*** **H* L H%**

It is virtually impossible to use **H*** on the postposed vocative; presumably it makes little sense for S to try to attract H's attention *after* making a point rather than before. If a **H*** is used on a postposed vocative, it has the flavor of a repair.

As a final example, work by Hirschberg and Litman (1987) on *cue phrases* lends support to our account of **L*** accent as excluding items from the predication of an utterance. Cue phrases are expressions such as *okay, but, now, anyway, by the way, in any case, that reminds me* that function to indicate discourse structure explicitly (Reichman 1985; Cohen 1984). Hirschberg and Litman analyzed the intonation of 100 instances of the word *now* in a corpus of recorded naturally occurring dialogues.[6] When *now* was used to signal (discourse) structural—rather than temporal—information, it often received a **L*** accent. In particular, in cases where *now* formed part of a larger intonational phrase, structural uses were either deaccented or accented with **L***. However, when deictic *now* formed part of a larger phrase, it received a **H*** or complex accent—never **L***. So, the communication of structural information correlates with the use of **L***, whereas the communication of temporal "content" correlates with the use of nonaccents. Of course, deictic *now* can have a **L*** accent in some cases that did not appear in the corpus—for example, if it is being questioned.

So, **L*** accents are used by S to exclude the accented item from the predication S intends to be added to H's mutual beliefs. There may be various reasons for and interpretations of this exclusion, including the use of **L*** in yes-no questions (where S requests H to make some predication), or to convey S's denial of some part of a previous predication, or to convey that the accented item already figures in what H currently believes to be mutually believed. Finally, **L*** is often used with items that have been independently analyzed as outside the predication of an utterance, such as greetings, vocatives, and cue phrases.

5.3 *The L+H Accents*

L+H accents are employed by S to convey the salience of some *scale* (defined here following Ward and Hirschberg 1985 as a partial ordering) linking the accented item to other items salient in H's mutual beliefs.

5.3.1 The L+H Accent* The interpretation of the **L*+H** pitch accent in the context of a **L** phrase accent and **H** boundary tone (**L*+H L H%**) has been intensively investigated by Ward and Hirschberg (1985, 1986). In the

1985 paper Ward and Hirschberg account for a large class of naturally occurring tokens in which this contour expresses uncertainty. They point out that, in all of their data, the contour is being used to convey uncertainty about a scale evoked in the discourse. For example, in (26) B expresses uncertainty about whether being a good badminton player provides relevant information about degree of clumsiness:

(26) A: Alan's such a klutz.
B: He's a good badminton player
L*+H L H%

And in (27) B conveys uncertainty about whether there are "degrees" of "taking out the garbage"—or whether it is an all-or-nothing phenomenon:

(27) A: Did you take out the garbage?
B: Sort of
L*+H L H%

In the 1986 paper Ward and Hirschberg address a second class of **L*+H L H%** uses, the "incredulous" readings. In (27), for instance, A might reply with (28) to convey that the proposed gradedness of garbage removal is unacceptable:

(28) A: Sort of
L*+H L H%

Ward and Hirschberg unify the "incredulous" and "uncertain" readings of **L*+H L H%** under the notion of "lack of speaker commitment" to the proposed scale (which they define as a partial ordering) or scalar value. A pilot phonetic study suggests that the difference between the two readings is conveyed by differences in pitch range and tempo.

We suspect that the contour interpretation that Ward and Hirschberg identify for **L*+H L H%** is more properly associated with the **L*+H** pitch accent rather than the entire contour. At least the "uncertainty" interpretation is still available when a **H** phrase accent is substituted for a **L** phrase accent, as in (29), in which a pet owner calls a missing and somewhat recalcitrant pet:

(29) Leo
L*+H H L%

And in the hypothetical (30), **L*+H** is paired with a **H** phrase accent and **H** boundary tone—with the same conveyance of uncertainty:

(30) A: We don't have any native speakers of German here. So let's work on Chinese.
B: Jurgen's from Germany
L*+H H H%

We believe that this account of the **L* + H** pitch accent can be recast in the present framework. We propose that S chooses a **L* + H** pitch accent to convey lack of predication and to evoke a scale. Together these can convey the impression of lack of speaker commitment described in Ward and Hirschberg 1986.

5.3.2 The L + H Accent* The meaning of the **L + H*** pitch accent is closely related to that of **L* + H**. Use of this accent also evokes a salient scale. However, S employs the **L + H*** accent to convey that the accented item—and not some alternative related item—should be mutually believed. The evocation of a salient scale plus predication can convey the effect of speaker commitment to the instantiation of the open expression with the accented item.

The most common use of **L + H*** in the data we have collected is to mark a correction or contrast. In such cases S substitutes a new scalar value for one previously proposed by S or by H—or for some alternative value available in the context. (31) occurred on a trip to Boston in December:

(31) A: It's awfully warm for January.
B: It's even warm for December
L + H* L H%

In (32) A and B were looking at the label of a Sambuca Romana bottle, which shows a man kissing the hand of a woman wearing a rather daring evening dress:

(32) A: I wonder if they're supposed to be married.
B: No, I don't think they're married.
If they were married, he wouldn't be kissing her hand
L + H* L H%

A class of cases discussed in Jackendoff 1972 is closely related. Jackendoff notes that the "background" information in dialogues like (33) has a distinctive intonation pattern:

(33) A: What about the beans? Who ate them?
B: Fred ate the beans
H* L L + H* L H%

The meaning assigned to this exchange is also "contrastive"—something like 'As for the beans, Fred ate them. As for the other food, other people may have eaten it'. Here, B's answer is felicitously produced in two phrases, *Fred* and *ate the beans*; the second, representing the background information, has a fall-rise pattern on *beans*. The phonological analysis of the pattern on *beans* is obscure from Jackendoff's description, but recent unpublished ex-

periments by Liberman and Pierrehumbert strongly suggest that it is **L+H* L H%**. Again, the "contrastive" interpretation can be accounted for in our framework as S's commitment to a particular instantiation of an open expression with an item chosen from a salient scale—here, a set of salient foods.

Example (34)—uttered while B was unpacking a new desk lamp—is less obviously interpreted as "contrastive." However, it is easily accommodated by our definition of the meaning of **L+H***.

(34) A: But how does it {the desk lamp} stand up?
B: Feel that base.
It weighs a ton
L+H* L+H* L H%

Here, its base "weighing a ton" is one of many possible means by which the lamp might stand up. B commits herself to this from the set of such means. In another example of **L+H*** a daughter calls her parents to invite them for dinner. Her mother consults with her father in (35):

(35) Mother: It's Raymond and Janet on the phone.
They want to know if we can come for dinner
L+H* L H%

Here, an invitation to dinner is implicitly related to a space of possible invitations, possible ways to spend the evening, or perhaps simply possible queries. The mother's use of **L+H*** conveys a strong sense of commitment to this accented item—which was interpreted by the daughter as indicating to her father that he should accept the invitation.

5.4 *The H+L Accents*

The **H+L** accents, like the **L+H** accents, are used by S to evoke a particular relationship between the accented items and H's mutual beliefs. **L+H** accents evoke a salient scale for the accented item. We propose that S uses **H+L** accents to indicate that support for the open expression's instantiation with the accented items should be inferable by H, from H's representation of the mutual beliefs. The inference can be direct or indirect, and it can be (and indeed usually is) pragmatic rather than logical in character. When using a **H*+L** accent, S appears to be making a predication in the same sense as when using **H***. **H*+L** thus differs from **H*** in conveying that H should locate an inference path supporting the predication. Items accented with **H*** might in principle be supported in the same way, but the support is not explicitly evoked by the tune.

We have collected only a few examples of the **H+L*** accent, and so we are less confident of its interpretation. In the examples we have, S seems to employ it to convey that the desired instantiation of an open expression is

itself among H's mutual beliefs. We conjecture that the basic meaning is the same as that for **H* + L**, except that **H + L*** does not make a predication. One reason for evoking support for an instantiation without making a predication would be (the claim) that the predication is already mutually believed. If this is correct, we should expect to find other contexts of use for this accent.

5.4.1 The H + L Accent* In some uses of **H* + L**, the inference path S wishes to evoke is so short that the accented items may alternatively be deaccented. In (36a) (due to Gregory Ward) S conveys both that the instantiation of 'I'm looking for someone with *x*' is particularly salient and that H should infer it from H's mutual beliefs. Perhaps, here, the relevant beliefs include the facts that S has mentioned H's credentials and an interview is in progress.

(36) I know you have great credentials
H* **H*** **H*** **L** **H%**

a. I'm looking for someone with just such credentials
H* **H*+L** **H*+L** **H*+L** **L** **L%**

b. I'm looking for someone with just such credentials
H* **L** **L%**

In (36b), however, S does not impart additional salience to H's credentials —which have already been made salient by S's **H*** accent in the previous utterance. Nor does S convey that H should look for an inference path between 'I'm looking for someone with *x*' and other of H's mutual beliefs.

However, the inference path may not always be so simple. In some cases **H* + L** accents can even be used discourse-initially. In the following (naturally occurring) exchange, C was a linguist whom colleagues A and B found particularly troublesome. A walked into B's office, where B was reading a circular advertising a linguistics position in Tasmania. B looked up and said,

(37) Let's nominate C for the Tasmanian job
H*+L **H*+L** **H*+L** **H*+L** **H*+L** **L** **L%**

Here, B invites A to consider not only the proposed nomination but also the path—C is obnoxious, obnoxious people should be got rid of, Tasmania is far away—by which it can be inferred.

The **H* + L** accent often has a pedagogical flavor. This is not surprising, since teaching involves pointing the student to inference relationships between old and new information. In assigning appropriate intonation to the synthesized speech for a computer-aided instruction system that teaches beginners how to use a screen editor (*TNT*), we found numerous cases where this accent was useful. In one case *TNT* introduces a "hint" key,

which provides learners with suggestions. Subsequent instructions to practice using this key were felicitously accented with **H* + L**:

(38) Hint gives you hints if you need help.
Hit the hint key
H* + L H* + L L L%

However, if the student were instructed to hit this key, say, without prior introduction to its function, **H* + L** accents would sound distinctly odd. A plain **H*** accent would be more appropriate.

The **H* + L** accent can also be used when reading instructions. For example, consider the series of **H* + L** accents in (39):

(39) Let's see
H* + L H* + L L L%
Put tab A into slot A
H* + L H* + L H* + L H* + L H* + L L L%
Turn the model over
H* + L H* + L H* + L L L%
Put tab B into slot B
H* + L H* + L H* + L H* + L H* + L L L%

Here, S is emphasizing the connections between each instruction and what S has already read or done.

Sometimes S uses **H* + L** accents when the inference path to be made salient is quite obscure to H—or even when S has no real belief that H will be able to discover it. In such cases **H* + L** sounds pretentious and annoying, as H is told it should be possible to infer something there is no obvious means of inferring.

In the examples given so far, a series of **H* + L** accents has been followed by a **L** phrase accent. When followed by a **H** phrase accent, the same pitch accent gives rise to a very distinctive pattern in which the voice trails out at a middle level, as in (40):

(40) Jimmy
H* + L H L%
Dinner
H* + L H L%

This pattern is often almost chanted and makes its first appearance in the literature as "calling" contour (Pike 1945).

Ladd (1978) points out that it is not really correct to call such contours vocative. He proposes instead that the pattern has a core meaning of "stylization" or shared convention. The contour is suitable even if the convention is a private one between individuals, as in (41), used to convey that H has forgotten his lunch yet again:

(41) Jacob
H*+L H L%
Your lunch
H*+L H L%

Finally, Ladd points out that this pattern is *in*appropriate for calling out in a real emergency, as in (42):

(42) #Fire
H*+L H L%

Our general interpretation of **H*+L** appears to account for Ladd's examples. (40) is most appropriate when Jimmy is expecting his dinner call, and (41) conveys that Jacob should be able to infer the reason that lunch is being brought to his attention. In addition, we cover some cases that Ladd misses by confining himself to "chanting" utterances. For instance, when one of us was pulled over by the police, the policeman said:

(43) Ma'am, your car inspection is overdue
H*+L H H%
I'll have to give you a summons.

In this case there is no past history of overdue car inspections. Rather, S is alluding to mutual beliefs established by the sticker on the windshield and the fact that H was pulled over.

5.4.2 ***H+L**** There is some difficulty in separating the meaning of **H+L*** from that of **H*+L**, because in many cases the phonological analysis in unclear. Both the **H*+L** and the **H+L*** accents create downstepping patterns (see section 3); they differ in whether there is an f_0 fall onto or after the accented syllable. If the accented syllables are very close together, the phonetic effect is much the same. However, the difference between the two is conspicuous if the accents are well separated or if the accent precedes a **H** phrase accent.

In the examples we have collected, **H+L*** is used to convey that the instantiation of the open expression is already present among H's mutual beliefs. Consider (44), in which S questions H's travel plans:

(44) It's inconceivable that we'll make that connection
H* H+L* L L%

H interpreted this utterance as conveying that H should already know this fact. In another instance one of us had a discussion with her mother-in-law in which they disagreed about why the baby had awakened in the middle of the night. In this discussion the mother-in-law advanced a mutually known fact as the correct explanation:

(45) She's teething
H* H+L* H L%

Though H+L* cannot be used in Ladd's "calling" sense, it is sometimes used with conventionalized expressions, as in (46):

(46) A: Janet, you've crashed Sweet again.
B: Oh darn it
H+L* H L%

The use of H+L* with this and other expletives has a sort of "redundant" effect—during part of its life, Sweet crashed very often. S does not use this accent with expletives when the situation occasioning the expletive is completely new. So, exchanges like (47) seem odd:

(47) A: I just heard we're not getting a pay raise this year. I don't understand—the company's doing so well!
B: # Oh darn it
H+L* H L%

In felicitous uses of H+L* with expletives, we might say that S is confirming a reaction previously recognized by H.

5.5 The Compositionality of Accent Meanings

In the description just proposed, the meanings of the starred tones are shared among the different accent types. When the starred tone is L (L*, L*+H, and H+L*), S does not convey that the instantiation of the open expression by the accented item should be added to H's mutual beliefs. For one of a variety of reasons—it may already be there, S may not be certain of its appropriateness, S may not wish or be able to predicate the open expression of the accented item—S does not intend to contribute this instantiation to H's mutual beliefs. However, when the starred tone is H (H*, L+H*, H*+L), S does intend to instantiate the open expression in H's mutual belief space. In addition, we note that items differing only in the location of the star have closely related meanings. L*+H and L+H* both evoke a salient scale. H*+L and H+L* both convey that H should be in a position to infer support for the instantiated expression—whether because it is already represented among H's mutual beliefs about S or because there is an inference path based on the mutual beliefs that supports the instantiation. These observations suggest that the meaning of each particular pitch accent may be derivable from the meanings of its constituent tones, plus some generalization about the interpretation of the star.

However, beyond the observations just made, we are not able to present such a decomposition as yet. The meanings of the two-tone accents all involve identifying a particular relationship between the (propositional

content of the) current utterance and H's mutual belief space: for the single-tone accents, no similar relationship is discernible. We have as yet no explanation of why complex tones should convey this additional meaning. Nor can we explain the difference between the **L + H** interpretation (identify a relevant scale) and the **H + L** interpretation (identify an inference path) in terms of composition from simple tones. This analysis we leave to further research. We also postpone the question of how phrases with mixed accent types are to be interpreted. Though we might propose a simple solution for mixed **H*** and **L*** accents—the former contributes to the predication, whereas the latter does not—combining, say, **L + H*** and **H* + L** in a single phrase will be more difficult to analyze.

6 *The Interpretation of Phrasal Tones*

Phrase accents have scope over entire intermediate phrases and may consist of either a high (**H**) or a low (**L**) tone (see section 2.3). These tones appear to indicate the presence or absence of an interpretive as well as a phonological boundary. A **H** phrase accent, for example, indicates that the current phrase is to be taken as forming part of a larger composite interpretive unit with the following phrase. A **L** phrasal tone emphasizes the separation of the current phrase from a subsequent phrase. Most of the support for this analysis comes from cases in which an intonational phrase is composed of several intermediate phrases—without intervening boundary tones. In the case of simpler intonational phrases—with but a single intermediate phrase—it is more difficult to separate the meaning of the phrase accent from the meaning of the boundary tone.

The use of a **H** phrase accent in listings appears to convey that the resulting list is intended to be exhaustive. For example, compare the use of the **H** tone in (48)–(49) with (50), in which a **L** phrase accent is used with the first item of the list:

(48) Do you want apple juice or orange juice
H* **H** **H*** **L** **L%**

(49) Do you want apple juice or orange juice
L* **H** **H*** **L** **L%**

(50) Do you want apple juice or orange juice
H* **L** **H*** **L** **L%**

We interpret this distinction in the following way. By using a **H** phrase accent in (48), S emphasizes that *apple juice* and *orange juice* form an entity, namely, the set of available juices; by using a **L** tone in (50), S emphasizes the separate status of each type of juice and thus does not evoke a larger interpretive entity.

Examples (48) and (49) appear to be virtually indistinguishable in meaning. We would suggest that the **L*** in (49) is marking a nonpredication because the predication is being deferred until the items in the list have been specified. From a functional point of view, then, the **L*** is reinforcing an interpretation that is independently conveyed by the phrase accent. Thus, **H*** and **L*** do not differ very much in their interpretation in exhaustive disjunctions.

Note that in both (48) and (50) an intermediate phrase boundary separates the disjuncts; thus, the distinction noted seems clearly due to the type of phrase accent and not to the presence or absence of a phrase break. It may be that failure to produce an intermediate phrase boundary between conjuncts simply leaves their interpretation as an exhaustive or partial set open to H's interpretation. However, the intuition that failure to produce a phrase boundary between the conjuncts might also lead to their interpretation as a single unit raises the question of how an utterance like (51) might differ from (48):

(51) Do you want apple juice or orange juice
L* **H*** **L** **L%**

To examine this question, we first note that (51) is somewhat unnatural, giving rise to the sense that the first mention of *juice* is somewhat unnecessary. That is, it is a reduced disjunction, *apple or orange juice*, that is actually desired. Since the distinction between disjunctions produced with or without internal phrase boundaries is clearest when scope ambiguities are possible, we might propose a more complex set of comparisons to tease apart the meaning of disjunctions with **H** phrasal tones from the meaning of disjunctions with no internal phrase boundaries. Compare (52) with (53) and (54):

(52) Do you want an apple or banana cake
L* **H*** **L** **L%**

(53) Do you want an apple or banana cake
H* **H** **H*** **L** **L%**

(54) Do you want an apple or banana cake
H* **L** **H*** **L** **L%**

Without an intermediate phrase boundary, as in (52), the disjunction is most plausibly interpreted as a modifier disjunction, *apple or banana*, modifying *cake*. However, with phrase boundaries, as in (53) and (54), it is most likely that both *an apple* and *banana cake* are being offered. In (53) these items are *all* that is being offered, whereas in (54) other foods may be available as well. So, we suggest that the presence or absence of a phrase boundary can influence the interpretation of the scope of disjunctions (and

conjunctions as well, as we discuss below). However, it is the type of phrase accent that conveys whether or not the resulting disjunction will be interpreted as exhaustive.

Choice of phrase accent can also influence the interpretation of relationships between conjoined clauses. It has long been noted (Schmerling 1976) that *and* is asymmetric. In some cases it can convey temporal, causal, or enablement relationships between conjoined clauses. We propose that choice of phrase accent can influence whether or not such an interpretation is conveyed. A **H** phrase accent can favor such an additional meaning; a **L** phrase accent does not. In (55), for example, a **H** tone favors the interpretation that George's ingestion of chicken soup caused his illness:

(55) George ate chicken soup and got sick
H* H* H* H H* H* L L%

In (56) the causal link—though still inferable—is not intonationally reinforced:

(56) George ate chicken soup and got sick
H* H* H* L H* H* L L%

In the more plausibly ambiguous (57) the role of the **H** tone in suggesting a causal link is more easily seen:

(57) I opened the door and the rain poured down
H* H* H H* H* L L%

In this example one seems clearly presented with a causal connection between S's action and a natural phenomenon, however implausible that might otherwise be. H is led to extended interpretations of the second conjunct—for example, it might be taken to mean 'the rain poured down on me'.

Similarly, the "implicit conditional" reading of conjunction is favored by **H** phrase accents, as in (58):

(58) Eat another cookie and I'll kill you
H* H H* L L%

7 *The Interpretation of Boundary Tones*

Boundary tones may also be **H** or **L** but have scope over the entire intonational phrase. As such, they appear to play a considerable role in the conveyance and perception of discourse segmentation. It is a common simplification in studies of discourse coherence to model discourses as sequences of declarative utterances in which the coherence of each new utterance is assessed with respect to those that precede it. Our findings

tend to support this model but suggest an additional simplifying generalization. As a first approximation, we propose that choice of boundary tone conveys whether the current intonational phrase is "forward-looking" or not—that is, whether it is to be interpreted with respect to some succeeding phrase or whether the direction of interpretation is unspecified. We propose that a **H** boundary tone indicates that S wishes H to interpret an utterance with particular attention to subsequent utterances. A **L** boundary tone does not convey such directionality.

Note that this proposal differs from the notion that **H** boundary tones signal "other-directed" utterances—those particularly designed to elicit a response. This claim derives particularly from the common occurrence of **H%** in yes-no questions. Though such questions may indeed be described as "other-directed" they are surely no more so than *wh*-questions, which typically are uttered with **L%**. And the "other-directed" generalization does *not* apply to other uses of **H%**.

Consider, for example, utterances bearing *continuation rise*—with a **L** phrase accent and **H** boundary tone—which need not be "other-directed." In a sequence like (59), for example, the **H** boundary tone on (59b) conveys that (59b) is to be interpreted with respect to a succeeding phrase, (59c)—not that (59b) itself is particularly intended to elicit a response:

(59) a. My new car manual is almost unreadable
L L%

b. It's quite annoying
L H%

c. I spent two hours figuring out how to use the jack
L L%

Now contrast (59) with (60). Use of the **H** boundary tone on (60a) tends to convey that (60a) is to be interpreted with respect to (60b):

(60) a. My new car manual is almost unreadable
L H%

b. It's quite annoying
L L%

c. I spent two hours figuring out how to use the jack
L L%

A consequence of these differences is that, whereas the referent of *it* in (60) is likely to be interpreted as *my new car manual*, the referent in (59) is likely to be understood to be 'my spending two hours figuring out how to use the jack'.

In these examples the "forward reference" signaled by a **H** boundary tone can be interpreted as 'this utterance will be completed by a subsequent

utterance'. Sequences of similar utterances can produce a similar effect, as in (61):

(61) a. George likes cake
L H%

b. He adores pie
L H%

c. He'll eat anything that's sweet and calorific
L L%

Both (61a) and (61b) are to be interpreted with respect to a succeeding utterance, (61c); in this case the sense that the first two utterances "are completed by" a third may be interpreted in Grosz and Sidner's (1986) terms as 'the intention underlying (61c) dominates those underlying both (61a) and (61b)', or in the terms of Cohen (1981), Mann and Thompson 1986, or Hobbs (1979) as '(61a) and (61b) provide evidence for (61c)'.

The **H** boundary tones used in yes-no question contours also convey "forward reference." Typically, this reference is cross-speaker. Any yes-no question-answer pair illustrates this phenomenon. For example, (62a)'s **H** boundary tone might also be glossed as '(62a) is to be completed by a subsequent phrase'—here, (62b):

(62) a. Does it snow a lot in New Jersey
H H%

b. It does this year
L L%

If the intentions underlying yes-no questions are something like 'make the status of some queried proposition P mutually believed among S and H'—and if a simple or cooperative response has a similar underlying intention as in (62)—then in Grosz and Sidner's (1986) terms, the satisfaction of the intention underlying (62b) contributes to the satisfaction of the intention underlying (62a). Thus, (62a) dominates (62b).

Note that, although **H** boundary tones contribute to the interpretation of intentional structure by signaling the existence of hierarchical relationships, the direction of the dominance relationship is not specified. In (59) and (61) the phrase ending with **H%** is to be dominated by a subsequent utterance. In (60) and (62), on the other hand, the **H%** phrase is to dominate a subsequent utterance. **H%** can also signal that Grosz and Sidner's (1986) satisfaction-precedence relationships hold between siblings, as in (63):

(63) a. Attach the jumper cables to the car that's running
L H%

b. Attach them to the car you want to start
L H%

c. Try the ignition
L H%

d. If you're lucky
L H%

e. you've started your car
L L%

The intentions underlying (63a–d) are dominated by that underlying (63e), and, in addition, each of the intentions (63a–c) satisfaction-precedes the next.

So, **H%** can be interpreted as signaling a hierarchical relationship between intentions underlying the current utterance and a subsequent one, although the former may either dominate or be dominated by the latter. In addition, **H%** may signal satisfaction-precedence relationships among intentions underlying sequences of utterances.

S uses a **L** boundary tone to convey that the current utterance may be interpreted without respect to subsequent utterances. Use of this tone throughout a discourse gives the impression that each intonational phrase has separate and equal status in the discourse. Use of **L%** in combination with other phrases ending with **H%** signals the hierarchical and satisfaction-precedence relations described above: in (59) and (60) phrases ending in **L%** "complete" phrases ending in **H%**; in our investigations to date, it does not appear that phrases ending in **H%** can signal this function. **L%** phrases can also represent siblings to phrases ending in **H%**, as the final element in lists. In a variation of (61), (intentions underlying) phrases (64a–c) are siblings dominated by (64d):

(64) a. George likes cake
L H%

b. He adores pie
L H%

c. He kills for chocolate mousse
L L%

d. He'll eat anything that's sweet and calorific
L L%

Another consequence of our account of the meaning conveyed by choice of boundary tone is that phrases with **H** boundary tones do not felicitously end discourse segments. In fact, violations of this generalization, such as (65), clearly convey that there is more that could or should be said:

(65) So, I guess there's just nothing more to say
L H%

In contrast, the L boundary tone, which does not convey such directionality of interpretation, can felicitously be used to begin discourse segments.

8 Discussion

In this paper we have presented the beginning of a compositional theory of the meaning of intonational contours. We propose that S chooses an intonational contour to convey relationships between (the propositional content of) the current utterance and previous and subsequent utterances—and between (the propositional content of) the current utterance and beliefs H believes to be mutually held. These relationships are conveyed compositionally via selection of pitch accent, phrase accent, and boundary tone. Pitch accents convey information about the status of discourse referents, modifiers, predicates, and relationships specified by accented lexical items. Phrase accents convey information about the relatedness of intermediate phrases—in particular, whether (the propositional content of) one intermediate phrase is to form part of a larger interpretive unit with another. Boundary tones convey information about the directionality of interpretation for the current intonational phrase—whether it is "forward-looking" or not. So, not only do different features of an intonational phrase convey different aspects of its meaning, but the meaning conveyed by each feature has scope over a different phonological domain. Together, pitch ccents, phrase accents, and boundary tones convey how H should interpret the current utterance structurally—with respect to previous and subsequent utterances—and with respect to what H believes to be mutually believed in the discourse.

Notes

1. We employ the distinction between attentional and intentional structure proposed in Grosz and Sidner 1986.
2. This correspondence between phonological and semantico-pragmatic domain of interpretation is suggested by work on Japanese phonology (Pierrehumbert and Beckman, 1988).
3. Grosz and Sidner propose a tripartite view of discourse structure: a *linguistic structure*, which is the text/speech itself; an *attentional structure*, which includes information about the relative salience of objects, properties, relations, and intentions at any point in the discourse; and an *intentional structure*, which relates *discourse segment purposes* (DSPs)—whose recognition is essential to a segment's achieving its intended effect—to one another. Each DSP contributes to the overall *discourse purpose* (DP) of the discourse. DPs and DSPs are intentions whose satisfaction represents the main purpose of a discourse or segment, for instance, "Intend that an agent believe some fact" or "Intend that an agent believe that one fact supports another." Although all DSPs by definition must contribute to the DP, DSPs are also related to one another in two ways. First, DSP1 is said to *contribute to* DSP2 when DSP1 provides part of the satisfaction of DSP2; in this case DSP2 is said to *dominate* DSP1. Second, DSP1 is said to *satisfaction-precede* DSP2 when-

ever DSP1 must be satisfied before DSP2. These relations thus impose two partial orderings on DSPs in a discourse: a dominance hierarchy and a satisfaction-precedence ordering.

4. The question of how an accent becomes associated with certain material is not yet well understood. For example, the general association of accent with components of NPs seems fairly clear: For example, stressing *DRESS* in *the girl in the red DRESS* may serve to focus the whole phrase, the PP, the smaller NP, or simply the N, *DRESS*—whereas stressing *RED* instead in the same phrase, *the girl in the RED dress*, can focus only the adjective. But it is not clear that the various focus possibilities in the first case are all realized identically—that the accented *DRESS* will in each case have the same prominence, for example. And similar constraints on the accenting of parts of a VP are even less well understood.
5. We are not yet prepared to propose a particular representation for intonational meaning, and so this depiction should be understood as metaphorical only. In particular, we do not intend that these open expressions represent the presupposition of an utterance, as previously suggested by Jackendoff (1972) and Wilson and Sperber (1979).
6. This corpus was recorded by Hirschberg and Pollack in 1982 from a Philadelphia radio call-in program, Harry Gross's "Speaking of Your Money" (Pollack, Hirschberg, and Webber 1982).

References

Anderson, M. D., J. B. Pierrehumbert, and M. Y. Liberman (1984). Synthesis by rule of English intonation patterns. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, San Diego, CA. Vol. 1.

Beckman, M., and J. Pierrehumbert (1986a). Intonational structure in Japanese and English. *Phonology Yearbook* 3, 15–70.

Beckman, M., and J. Pierrehumbert (1986b). Japanese prosodic phrasing and intonation synthesis. In *Proceedings of the Twenty-fourth Annual Meeting*, Association for Computational Linguistics, New York, NY.

Bolinger, D. (1958). A theory of pitch accent in English. *Word* 14, 109–149.

Brazil, D., M. Coulthard, and C. Johns (1980). *Discourse intonation and language teaching*. London: Longman.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research* 4, 75–87.

Clark, H. H., and C. R. Marhsll (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. Webber, and I. A. Sag, eds., *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Cohen, R. (1981). Investigation of processing strategies for the structural analysis of arguments. In *Proceedings of the Nineteenth Annual Meeting*, Association for Computational Linguistic. Stanford, CA.

Cohen, R. (1984). A computational theory of the function of clue words in argument understanding. In *Proceedings of the Tenth International Conference on Computational Linguistics*, Stanford, CA.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Grosz, B., and C. Sidner (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics* 12, 175–204.

Gussenhoven, C. (1983). On the grammar and semantics of sentence accents. Dordrecht, Holland: Foris.

Hirschberg, J., and D. Litman (1987). Now let's talk about 'now': Identifying cue phrases intonationally. In *Proceedings of the Twenty-fifth Annual Meeting*, Association for Computational Linguistics, Stanford, CA.

Hirschberg, J., and J. Pierrehumbert (1986). The intonational structuring of discourse. In *Proceedings of the Twenty-fourth Annual Meeting*, Association for Computational Linguistics, New York, NY.

Hobbs, J. (1979). Coherence and coreference. *Cognitive Science* 3, 67–90.

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Joshi, A. K. (1982). The role of mutual beliefs in question-answer systems. In N. Smith, ed., *Mutual knowledge*. New York: Academic Press.

Ladd, D. R. (1978). Stylized intonation. *Language* 54, 517–540.

Ladd, D. R. (1980). *The structure of intonational meaning*. Bloomington, IN: Indiana University Press.

Ladd, D. R. (1983). Phonological features of intonational peaks. *Language* 59, 721–759.

Ladd, D. R., K. Silverman, F. Tolkmitt, G. Bergmann, and K. Scherer (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America* 78, 435–444.

Liberman, M. (1975). The intonational system of English. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA. Published by Garland Press, New York, 1979.

Liberman, M., and J. Pierrehumbert (1984). Intonational invariants under changes in pitch range and length. In M. Aronoff and R. Oehrle, eds., *Language sound structure*. Cambridge, MA: MIT Press.

Liberman, M., and I. A. Sag (1974). Prosodic form and discourse function. In *Papers from the Tenth Regional Meeting*, Chicago Linguistic Society, Chicago, IL.

Mann, W. C., and S. A. Thompson (1986). Relational propositions in discourse. *Discourse Processes* 9, 57–90.

O'Connor, J. D., and G. F. Arnold (1961). Intonation of colloquial English. London: Longmans.

Pierrehumbert, J. B. (1980). The phonology and phonetics of English intonation. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Pierrehumbert, J. B., and M. Beckman (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J. B., and S. Steele (1987). How many rise-fall-rise contours? In *Proceedings of the Eleventh Meeting*, International Congress of Phonetic Sciences, Tallinn.

Pierrehumbert, J. B., and S. Steele (in press). Categories of tonal alignment in English. *Phonetica*.

Pike, K. (1945). The intonation of American English. Ann Arbor, MI: University of Michigan Press.

Pollack, M. E., J. Hirschberg, and B. Webber (1982). User participation in the reasoning processes of expert systems. MS-CIS-82-9. University of Pennsylvania Philadelphia, PA. A shorter version appears in the *Proceedings of the National Conference*, Association for Artificial Intelligence, Pittsburgh, PA.

Reichman, R. (1985). *Getting computers to talk like you and me: Discourse context, focus, and semantics*. Cambridge, MA: MIT Press.

Sag, I. A., and M. Liberman (1975). The intonational disambiguation of indirect speech acts. In *Papers from the Eleventh Regional Meeting*, Chicago Linguistic Society, Chicago, IL.

Schegloff, E. A. (1979). The relevance of repair to syntax-for-conversation. In T. Givon, ed., *Syntax and semantics 12: Discourse and syntax*. New York: Academic Press.

Schmerling, S. (1976). *Aspects of English sentence stress*. Austin, TX: University of Texas Press.

Selkirk, E. O. (1984). *Phonology and syntax*. Cambridge, MA: MIT Press.

Silverman, K. (1987). Natural prosody for synthetic speech. Doctoral dissertation, Cambridge University, Cambridge.

Ward, G., and J. Hirschberg (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61, 747–776.

Ward, G., and J. Hirschberg (1986). Reconciling uncertainty with incredulity: A unified account of the L*+H L H% intonational contour. Paper presented at the Linguistic Society of America Annual Meeting.

Wilson, D., and D. Sperber (1979). Ordered entailments: An alternative to presuppositional theories. In C.-K. Oh and D. A. Dinneen, eds., *Syntax and semantics 11: Presupposition*. New York: Academic Press.