Pragmatics and Intonation

Julia Hirschberg
AT&T Labs — Research
A257 — Shannon Laboratory
Florham Park, NJ 07932-0971
julia@research.att.com
http://www.research.att.com/ julia/index.html

1 Introduction

There is a long tradition of research on the role of prosodic variation in the interpretation of a wide variety of linguistic phenomena (Ladd, 1980; Ladd, 1996; Bolinger, 1986; Bolinger, 1989). Whether a speaker says (where '|' is read as a prosodic boundary and capitals denote emphasis) *John only introduced MARY to Sue* or *John only introduced Mary to SUE*; *Bill doesn't drink | because he's unhappy* or *Bill doesn't drink because he's unhappy* **can**, in the appropriate context, favor different interpretations of the same sentence. Since the interpretation of such intonational variations is indeed dependent upon contextual factors, we will define intonational "meaning" as essentially pragmatic in nature.

In this chapter, we will provide an overview of various types of intonational variation and the interpretations such variation has been found to induce. While the very large literature on intonational meaning from the linguistics, computational linguistics, speech, and psycholinguistic communities makes it impossible to provide an exhaustive list of relevant research efforts on the topic, examples of such work will be provided in each section. In Section 2, we will first describe the components of intonational variation that will be addressed in this chapter, employing as a framework for intonational description — the ToBI system for representing the intonation of standard American English. In Section 3, we will survey some of the ways intonation can influence the interpretation of syntactic phenomena, such as attachment. In Section 4 we will examine intonational variation and semantic phenomena such as scope ambiguity and association with focus. In Section 5, we will turn to discourse-level phenomena, including the interpretation of pronouns, the intonational correlates of several types of information status, the relationship between intonational variation and discourse structure, and the role of intonational variation in the interpretation of different sorts of speech acts. A final section will point to future areas of research in the pragmatics of intonation.

2 Intonation: Its Parts and Representations

To discuss prosodic variation usefully, one must choose a framework of intonational description within which to specify the dimensions of variation. The intonational model we will assume below below is the ToBI model for describing the intonation of standard American English (Pitrelli, Beckman, and Hirschberg, 1994; Silverman et al., 1992).[1] The ToBI system consists of annotations at four, time-linked levels of analysis: an ORTHOGRAPHIC TIER of time-aligned words; a BREAK INDEX TIER indicating degrees of junction between words, from *0* 'no word boundary' to *4* 'full INTONATIONAL PHRASE boundary', which derives from Price et al. (1990); a TONAL TIER, where PITCH ACCENTS, PHRASE ACCENTS and BOUNDARY TONES describing targets in the FUNDAMENTAL FREQUENCY (f0) define intonational phrases, following Pierrehumbert's (1980) scheme for describing American English, with some modifications; and a MISCELLANEOUS TIER, in which phenomena such as disfluencies may be optionally marked.

Break indices define two levels of phrasing: minor or INTERMEDIATE PHRASE (in Pierrehumbert's terms) (level 3); and major or INTONATIONAL PHRASE (level 4), with an associated tonal tier that describes the phrase accents and boundary tones for each level. Level 4 phrases consist of one or more level 3 phrases, plus a high or low boundary tone (**H%** or **L%**) at the right edge of the phrase. Level 3 phrases consist of one or more pitch accents, aligned with the stressed syllable of lexical items, plus a PHRASE ACCENT, which also may be high (**H-**) or low (**L-**). A standard declarative contour, for example, ends in a low phrase accent and low boundary tone, and is represented by **L-L%**; a standard yes-no-question contour ends in **H-H%**. These are illustrated in Figures 1 and 2, respectively.[2]

Differences among ToBI break indices can be associated with variation in f0, PHRASE-FINAL LENGTHENING (a lengthening of the syllable preceding the juncture point), glottalization ("creaky voice") over the last syllable or syllables preceding the break, and some amount of pause. Higher number indices tend to be assigned where there is more evidence of these phenomena. Phrasal tone differences are reflected in differences in f0 target.

Pitch accents render items intonationally prominent. This prominence can be achieved via different tone targets, as well as differences in f0 height, to convey different messages (Terken, 1997; Campbell and Beckman, 1997). So, items may be accented or (DEACCENTED (Ladd, 1979b)) and, if accented, may bear different tones, or different degrees of prominence, with respect to other accents. In addition to f0 excursions, accented words are usually louder and longer than their unaccented counterparts. In addition to variation in type, accents may have different levels of prominence; i.e., one accent may be perceived as more prominent than another due to variation in f0 height or amplitude, or to location in the intonational phrase. Listeners usually perceive the last accented item in a phrase as the most prominent in English. This most prominent accent in an intermediate phrase is called the phrase's NUCLEAR ACCENT or NUCLEAR STRESS. Constraints on nuclear (sometimes termed sentence) stress are discussed by many authors including Cutler and Foss (1977), Erteschik-Shir and Lappin (1983), Schmerling (1976; 1974), and Bardovi-Harlig (1983b). Despite

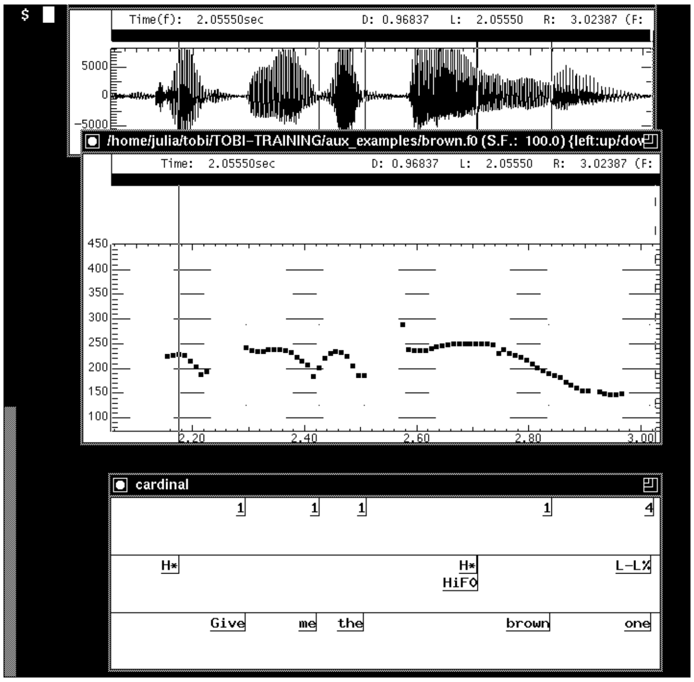


Figure 1: A **H* L- L%** Contour

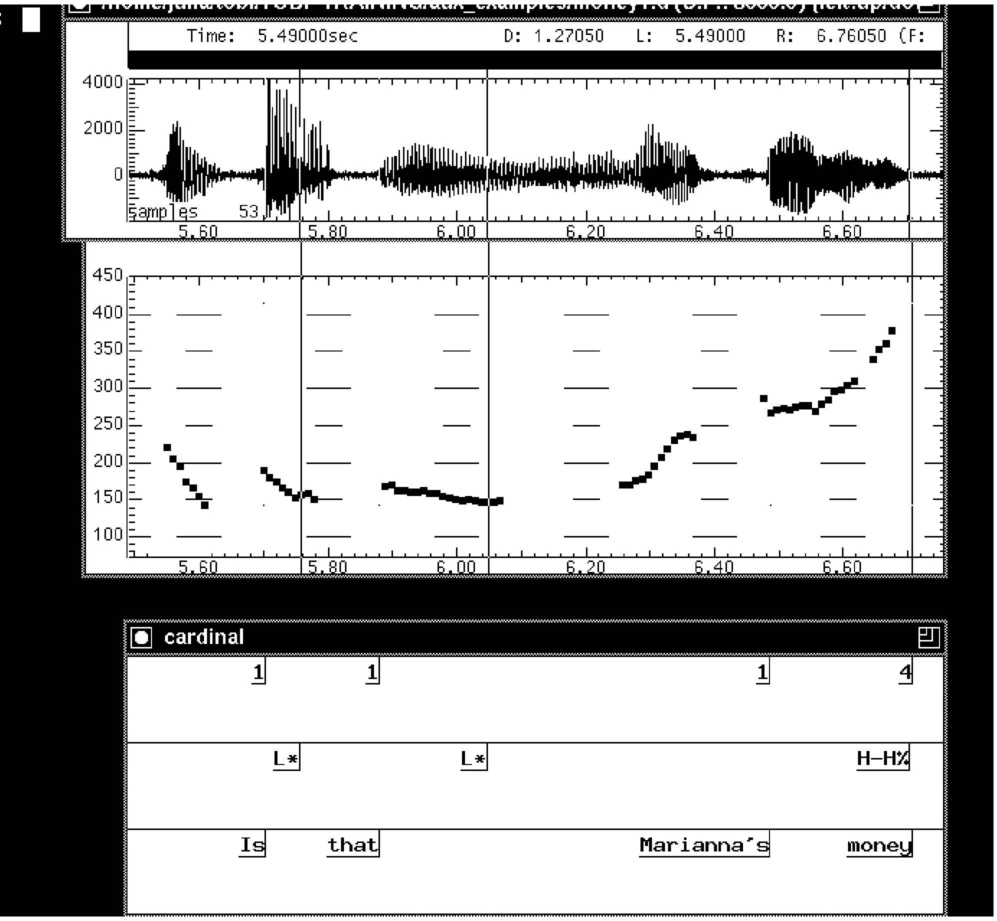


Figure 2: A **L* H- H%** Contour

Bolinger's (1972b) seminal article on the unpredictability of accent, attempts to do so from related features of the uttered text continue, especially for purposes of assigning accent in text-to-speech systems e.g. (Altenberg, 1987; Hirschberg, 1993; Veilleux, 1994).

Five types of pitch accent are distinguished in the ToBI scheme for American English: two simple accents **H*** and **L***, and three complex ones, **L*+H**, **L+H***, and **H+!H***. As in Pierrehumbert's system,[3] the asterisk indicates which tone is aligned with the stressed syllable of the word bearing a complex accent. Differences in accent type convey differences in meaning when interpreted in conjunction with differences in the discourse context and variation in other acoustic properties of the utterance. The **H*** accent is the most common accent in American English. It is modeled as a simple peak in the f0 contour, as illustrated in Figure 1 above; this peak is aligned with the word's stressable syllable.

H* accents are typically found in standard declarative utterances; they are commonly used to convey that the accented item should be treated as NEW information in the discourse, and is part of what is being asserted in an utterance (Pierrehumbert and Hirschberg, 1990). **L*** accents are modeled as valleys in the f0, as shown in Figure 2 above.

These accents have been broadly characterized as conveying that the accented item should be treated as **salient** but not part of what is being asserted (Pierrehumbert and Hirschberg, 1990). As such, they typically characterize prominent items in *yes-no* question contours. In addition to this use, they are often employed to make initial prepositions or adverbs prominent or to mark DISCOURSE readings of CUE PHRASES (see Section 5.3 below). **L+H*** accents can be used to produce a pronounced "contrastive" effect, as in (1a).

(1) The Smiths aren't inviting anybody important.

a. They invited **L+H*** Loraine.

b. They invited **L*+H** Loraine.

This complex accent, where the high tone is aligned with the stressed syllable and the f0 rise is thus rapid, can serve to emphatically contradict the initial claim that Loraine is unimportant and is illustrated in Figure 3. A similarly shaped accent with slightly but crucially different alignment, the **L*+H** accent, can convey still other distinctions. For example, **L*+H** pitch accent on *Loraine* in (1b), where the low tones is aligned with the stressed syllable, can convey uncertainty about whether or not Loraine is an important person. This type of accent is shown in Figure 4. And **H+!H*** accents, realized as a fall onto the stressed syllable, are associated with some implied sense of familiarity with the mentioned item. An example of a felicitous use of **H+!H*** is the "reminding" case in (2) and the accent is illustrated in Figure 5.

(2) A: No German has ever won the Luce Prize.
B: **H+!H*** Joachim's from Germany.

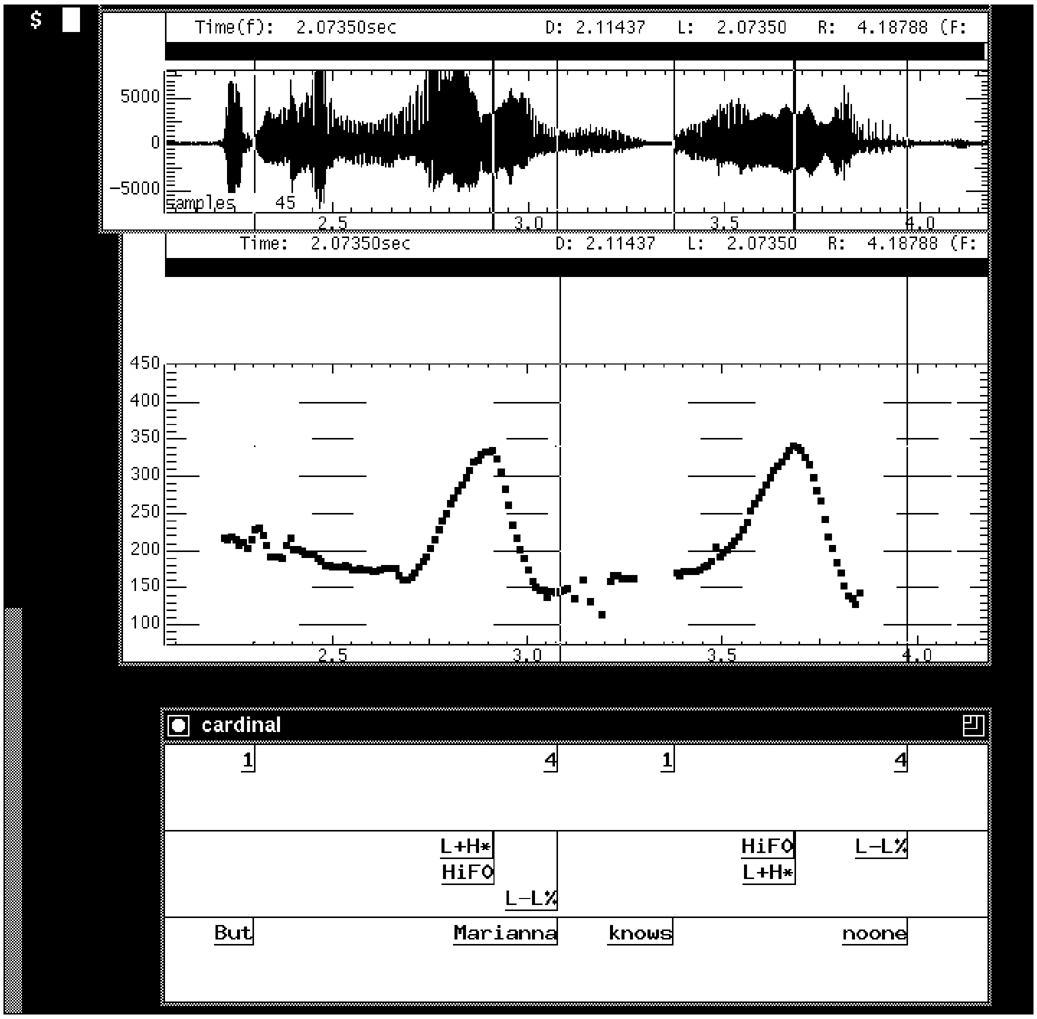


Figure 3: A **L+H*** Pitch Accent

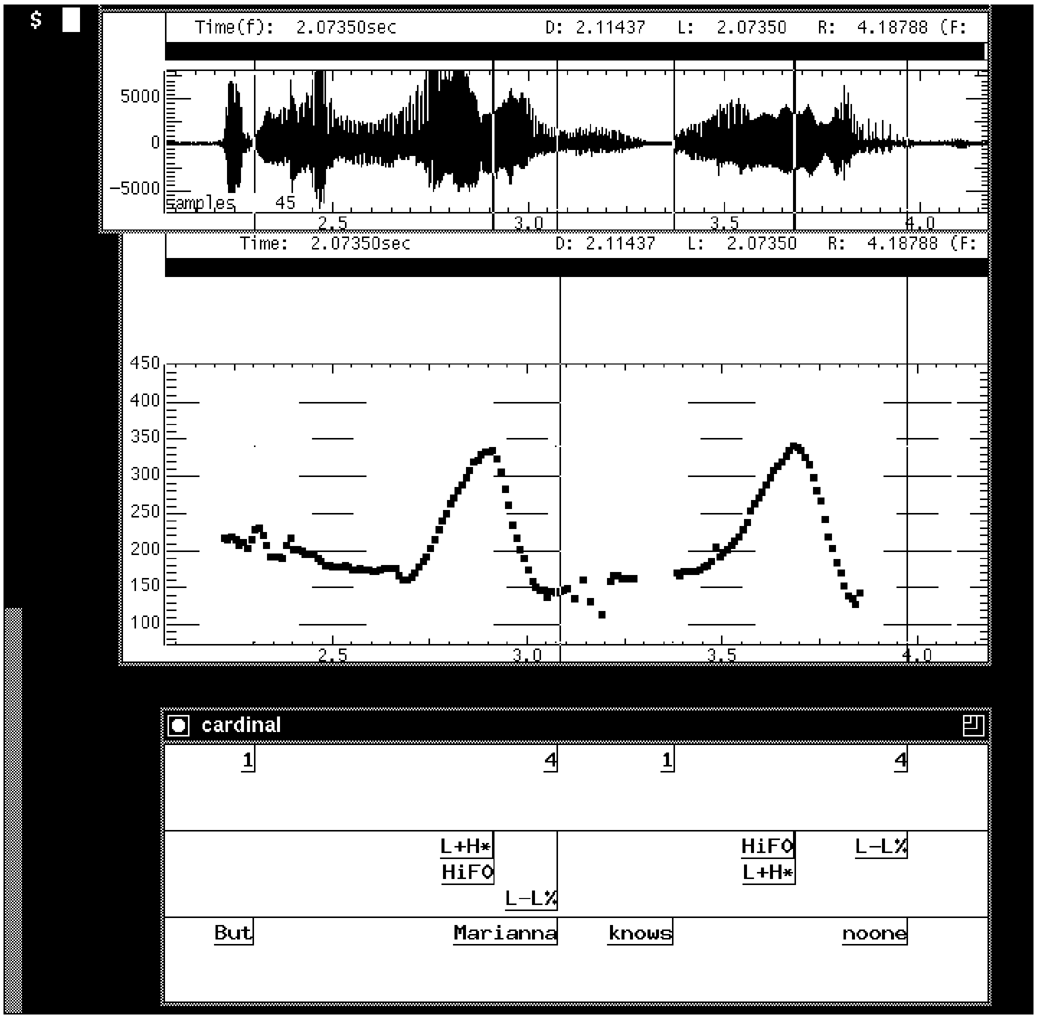


Figure 4: A **L*+H** Pitch Accent

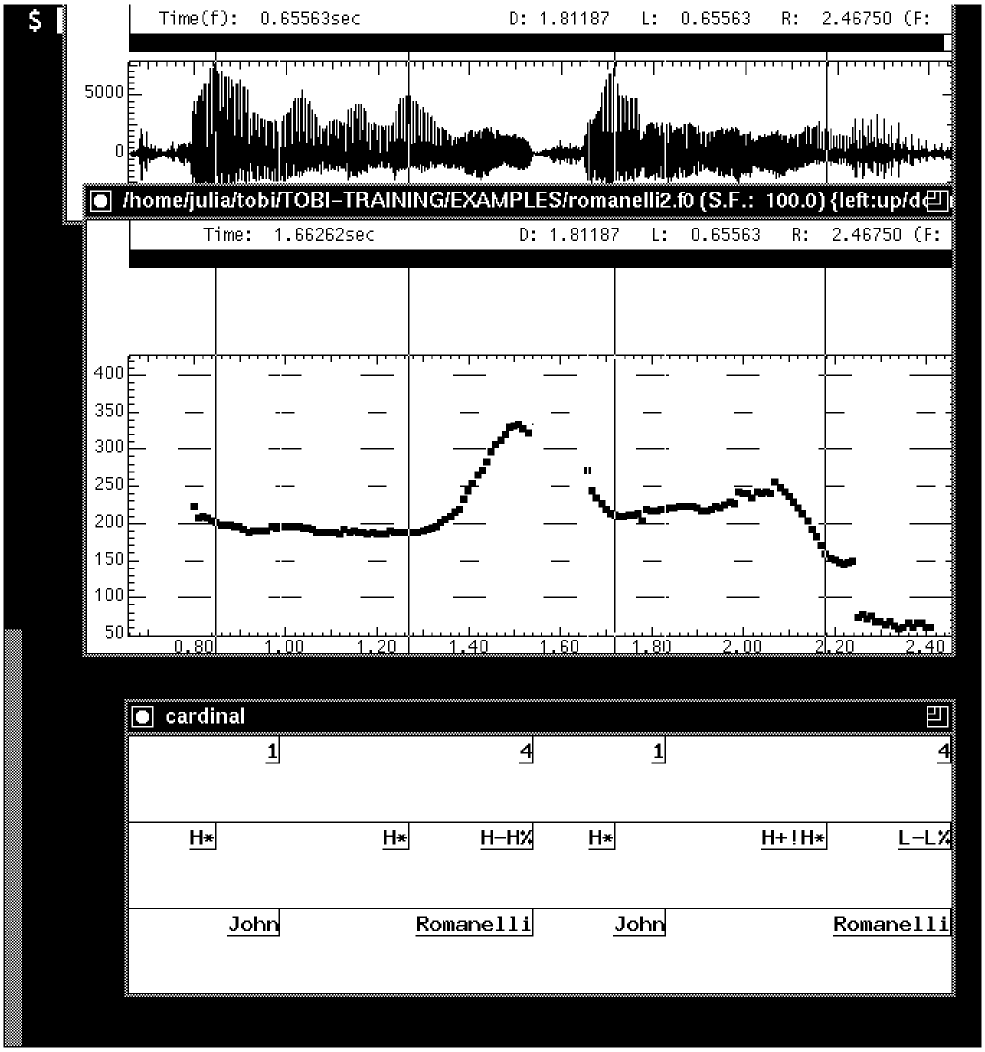


Figure 5: A **H+!H*** Pitch Accent

By way of summary, Table 1 provides a schematic representation of the possible contours in Standard American English, in the ToBI system.

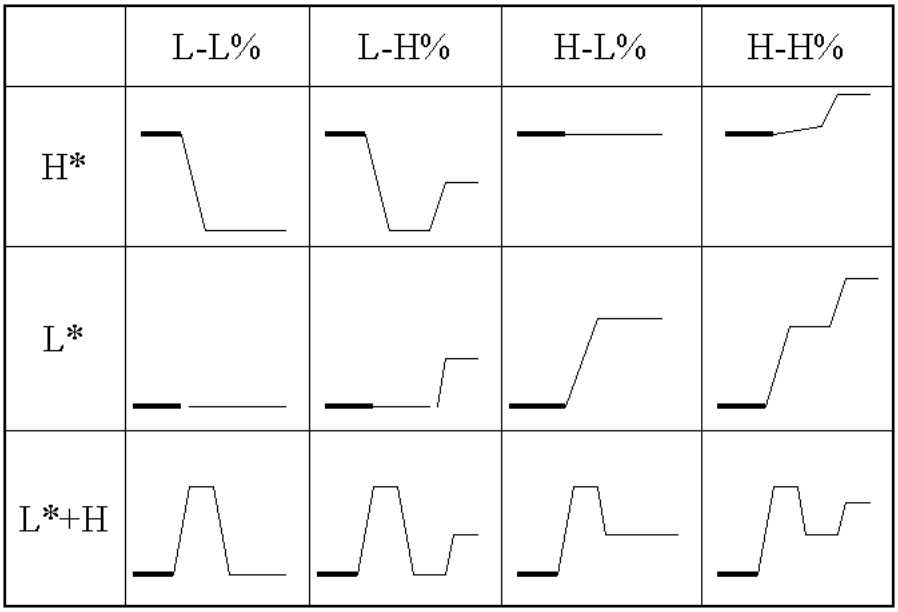
3 Intonation in the Interpretation of Syntactic Phenomena

There has been much interest among theorists over the years in defining a mapping between prosody and syntax (Downing, 1970; Bresnan, 1971; Selkirk, 1984; Cooper and Paccia-Cooper, 1980; Dirksen and Quene, 1993; Prevost and Steedman, 1994; Boula de Mareüil and d'Alessandro, 1998). Intuitively, prosodic phrases, whether intermediate or intonational, divide an utterance into meaningful "chunks" of information (Bolinger, 1989); the greater the perceived phrasing juncture, the greater the discontinuity between segments or constituents. While many researchers have sought to identify simple syntactic constraints on phrase location (Crystal, 1969; Cooper and Paccia-Cooper, 1980; Selkirk, 1984; Croft, 1995), especially for parsing (Marcus and Hindle, 1990; Steedman, 1991; Oehrle, 1991; Abney, 1995), more empirical approaches have focussed upon discovering the circumstances under which one sort of phrasing of some syntactic phenomenon will be favored over another by speakers and perhaps differently interpreted by hearers. Corpus-based studies (Altenberg, 1987; Bachenko and Fitzpatrick, 1990; Ostendorf and Veilleux, 1994; Hirschberg and Prieto, 1996; Fujio, Sagisaka, and Higuchi, 1997) and laboratory experiments (Grosjean, Grosjean, and Lane, 1979; Wales and Toner, 1979; Gee and Grosjean, 1983; Price et al., 1990; Beach, 1991; Hirschberg and Avesani, 1997) have variously found that the discontinuity indicated by a phrase boundary may serve to favor various differences in the interpretation of syntactic attachment ambiguity, for phenomena such as prepositional phrases, relative clauses, adverbial modifiers. Moreover, it has been found that the presence or absence of a phrase boundary can distinguish prepositions from particles and can indicate the scope of modifiers in conjoined phrases. Some examples are found in (3)-(11), where boundaries are again marked by '|':

(3) VP-attachment: Anna frightened the woman | with the gun.
[Anna held the gun]
Anna frightened | the woman with the gun.
[the woman held the gun]

(4) Mary knows many languages you know.
[Complementizer: Mary knows many languages that you also know]
Mary knows many languages | you know.
[Parenthetical: as you are aware, Mary knows many languages]

(5) The animal that usually fights the lion is missing.
[the lion's normal opponent is missing]



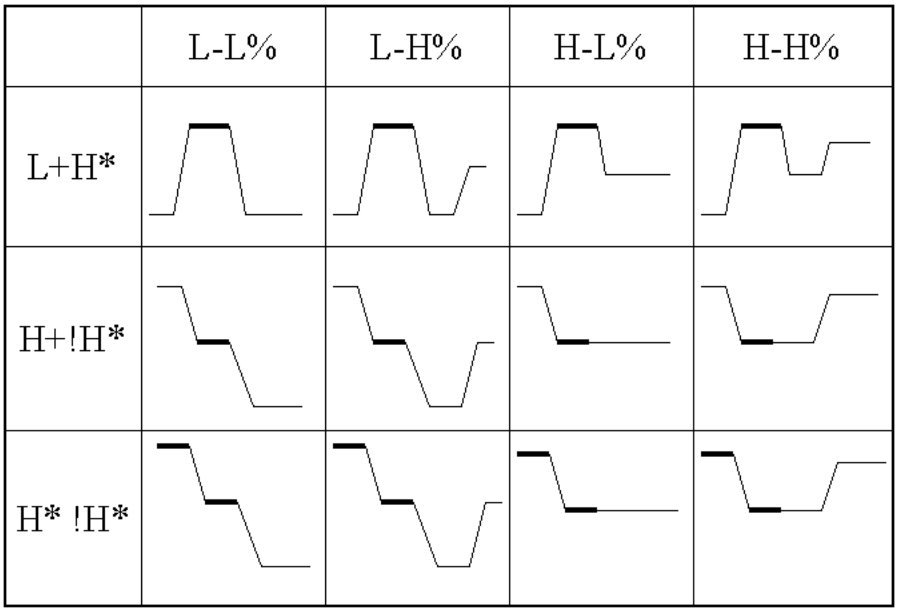


Table 1: ToBI Contours for Standard American English

The animal that usually fights | the lion | is missing.
[Appositive: the lion is missing]

(6) My brother who is a writer needs a new job.
[Restrictive relative clause: I have at least one other brother but I am not speaking of him]
My brother | who is a writer | needs a new job.
[Non-restrictive relative clause: I may or may not have other brothers]

(7) John laughed | at the party.
[Preposition: John laughed while at the party]
John laughed at | the party.
[Particle: John ridiculed the party]

(8) If you need me | when you get there call me.
[Attachment to antecedent clause VP: if you need me when you arrive, call me]
If you need me when you get there | call me.
[Attachment to main clause VP:if you need me, call me when you arrive]

(9) This collar is dangerous to younger | dogs and cats.
[Conjunction modification: the collar may be dangerous to younger dogs and younger cats]
This collar is dangerous to younger dogs | and cats.
[Single conjunct modification: the collar may be dangerous to younger dogs and all cats]

(10) Stir in rice wine | and seasonings.
[Compound noun: stir in two ingredients]
Stir in rice | wine | and seasonings.
[List interpretation: stir in three ingredients]

(11) We only suspected | they all knew that a burglary had been committed.
[Simple complement: we only suspected that they all knew that a burglary had been committed]
we only suspected | they all knew | that a burglary had been committed.
[Parenthetical: they all knew that we only suspected that a burglary had been committed]

Prosodic variation other than phrasing can also influence disambiguation of syntactic ambiguity. For example, range and rate can also distinguish phenomena such as parenthetical phrases from others (Kutik, Cooper, and Boyce, 1983; Grosz and Hirschberg, 1992): parentheticals like that in (11) are generally uttered in a compressed pitch range and with a faster speaking rate than other phrases. And the location of pitch accent can cue the right node raising reading of the sentence uttered in (11), as in (12) (Marcus and Hindle, 1990).

(12) WE only SUSPECTED | THEY all KNEW | that a BURGLARY had been committed.
[we suspected but they in fact knew that a burglary had been committed]

Pitch accent location is also a well known factor in conveying the structure of complex nominals (Liberman and Sproat, 1992; Sproat, 1994) and in distinguishing among part-of-speech ambiguities, as evident from the examples in (13) and (14):

(13) GERMAN teachers
German TEACHERS

(14) LEAVE in the LIMO
leave IN the REFERENCE

In (13), accent on the modifier or head signals the different interpretations: 'teachers of German' vs. 'teachers who are German'. And differences in accent location distinguish prepositions from VERBAL PREPOSITIONS as in (14). But note that prepositions may also be accented, to convey focus or contrast, as illustrated in (15).

(15) I didn't shoot AT him, I shot PAST him.

So, the relationship between accent and part-of-speech is also dependent upon context.

And, while intonational variation **can** serve all these functions, evidence that it does so reliably is mixed (Wales and Toner, 1979; Cooper and Paccia-Cooper, 1980; Nespor and Vogel, 1983; Schafer et al., 2000). Speakers routinely violate all of the distinctions illustrated above, perhaps because they do not recognize the potential ambiguity of their utterances or because context disambiguates. Even when explicitly asked to disambiguate, they may choose different methods of disambiguation.

4 Intonation in the Interpretation of Semantic Phenomena

There is a long and diverse tradition of research on the role of accent in the interpretation of semantic phenomena, centering around the interpretation of FOCUSSED constituents (Lakoff, 1971b; Schmerling, 1971b; Jackendoff, 1972; Ball and Prince, 1977; Wilson and Sperber, 1979; Enkvist, 1979; Gussenhoven, 1983; Culicover and Rochemont, 1983; Rooth, 1985; Rochemont and Culicover, 1990; Rooth, 1992; Horne, 1985; Horne, 1987; Baart, 1987; Dirksen, 1992; Zacharski, 1992; Birch and Clifton, 1995). Changing the location of nuclear stress in an utterance can alter the interpretation of the utterance by altering its perceived focus. An utterance's focus may be identified by asking 'To what question(s) is the utterance with this specified accent pattern a felicitous answer?' (Halliday, 1967; Eady and Cooper, 1986). For example, (16b) is a felicitous response to the question *Whom did John introduce to Sue?*, while (16c) is an appropriate response to the question *To whom did John introduce Mary?* In each case the focussed information is the information being requested, and is the most prominent information in the utterance.

(16) a. John only introduced Mary to Sue.

b. John only introduced MARY to Sue.

c. John only introduced Mary to SUE.

In (16b), Mary is the only person John introduced to Sue; in (16c), Sue is the only person John introduced Mary to. (16b) is false if John introduced Bill, as well as Mary, to Sue; (16c) is false if John introduced Mary to Bill, as well as to Sue. This variation in focus takes on an added dimension when FOCUS-SENSITIVE OPERATORS, such as *only* are present (Halliday, 1967; Jackendoff, 1972; Rooth, 1985; Sgall, Hajičova, and Panevová, 1986; Partee, 1991; Rooth, 1992; Vallduvì, 1999; Selkirk, 1995; Schwarzschild, 1999; Büring, 1999). In (16), the FOCUS-SENSITIVE OPERATOR *only* interacts with the intonational prominence of pitch accents to produce the different interpretations of the sentence discussed above. Other such operators include other quantifiers (*all*, *most*, *some*), adverbs of quantification (*sometimes*, *most often*), modals (*must*), emotive factives/attitude verbs (*It's odd that*), and counterfactuals. With prominence on *night*, for example, (17a) is felicitous, with the meaning 'it is at night that most ships pass through the lock'.

(17) a. Most ships pass through the lock at night.

b. When do ships go through the lock?
Most ships pass through the lock at NIGHT.

c. What do ships do at night?
Most ships pass through the LOCK at night.

However, with prominence on *lock*, the same sentence becomes an felicitous answer to a different question, as illustrated in (17b): Passing through the lock is what most ships do at night. Temporal quantification behaves similarly, as illustrated in (18). Other temporal quantifiers include *frequently*, *rarely*, *sometimes*, *occasionally*, and so on.

(18) a. Londoners *most often* go to Brighton.

b. Who goes to Brighton?
LONDONERS most often go to Brighton.

c. Where do Londoners go on vacation?
Londoners most often go to BRIGHTON.

The well-known example in (19a) was originally observed on a sign on a British train by Halliday (1967), who was startled to learn that every train rider was commanded to carry a dog, under the reading induced by (19b).

(19) a. Dogs must be carried.

b. DOGS must be carried.

c. Dogs must be CARRIED.

A more likely interpretation of the sentence in the railway context would be favored by the accent pattern represented by (19c): If you bring a dog on the train, then you must carry it. Other modals which associate with focus in a similar way include *can*, *should*, *may*.

Other focus-sensitive operators that also appear to identify the scope of the operator within the utterance are illustrated in (20)-(21):

(20) a. It's ODD that Clyde married Bertha.
 b. It's odd that CLYDE married Bertha.
 c. It's odd that Clyde MARRIED Bertha.
 d. It's odd that Clyde married BERTHA.

Depending upon whether *odd*, the operator itself, or one of its potential foci (*Clyde*, *married*, or *Bertha*) bears nuclear stress, what is "odd" may vary considerably. The entire proposition that Clyde married Bertha is odd. The fact that it was Clyde and not someone else who married Bertha is odd. What is odd if that what Clyde did with respect to Bertha was to marry her. Or it is the fact that the person Clyde chose to marry was indeed Bertha that is strange. And in (21), a listener would be likely to draw very different inferences depending upon the speaker's location of nuclear stress.

(21) a. This time HARRY didn't cause our defeat.
 b. This time Harry didn't CAUSE our defeat.
 c. This time Harry didn't cause our DEFEAT.

Someone else caused our defeat, not Harry (21a); Harry didn't actually cause our defeat though he may have, e.g., contributed to it (21b); Harry didn't cause our defeat but rather he caused something else (21c).

Although most research on the role of intonation in semantic interpretation has concentrated on pitch accent variation, variation in phrasing can also change the semantic interpretation of an utterance, again though with considerable variation in performance. For example, the interpretation of negation in a sentence like (22) is likely to vary, depending upon whether it is uttered as one phrase (22a) or two (22b).

(22) a. Bill doesn't drink because he's unhappy.
 b. Bill doesn't drink | because he's unhappy.

In (22a) the negative has wide scope: Bill does indeed drink — but the cause of his drinking is not his unhappiness. In (22b), it has narrow scope: Bill's unhappiness has lead him **not** to drink. However, like other interpretations that may be favored by intonational variation, if context itself can disambiguate a potentially ambiguous sentence, speakers sometimes produce intonational phrasings that do not obey these likelihoods. For example, an utterance of *Bill doesn't drink because he's unhappy* as a single phrase may be interpreted with the narrow scope of negation as well as the wide (Hirschberg and Avesani, 2000); interesting, *Bill doesn't drink | because he's unhappy* is less likely to be interpreted with wide scope negation. Such cases where a particular intonational pattern may be interpreted in several ways — but its contrast is less likely to — give rise to the notion of "neutral" intonation, a notion whose evidence is probably more persuasive for phrasing variations than for accent variation.

5 Intonation in the Interpretation of Discourse Phenomena

Intonational variation has been much studied in its role in the interpretation of numerous discourse phenomena. Pronouns have been found to be interpreted differently depending upon whether they are prominent or not, in varying contexts. Different categories of information status, such as THEME/RHEME distinctions, GIVEN/NEW status, and contrast, are believed to be intonationally markable (Schmerling, 1975b; Bardovi-Harlig, 1983a; Brown, 1983; Gundel, 1978; Lehman, 1977; Fuchs, 1980; Chafe, 1976; Nooteboom and Terken, 1982; Fuchs, 1984; Terken, 1984; Terken, 1985; Terken and Nooteboom, 1987; Fowler and Housum, 1987; Horne, 1991a; Horne, 1991b; Allerton and Cruttenden, 1979; Kruyt, 1985; Cahn, 1998; Terken and Hirschberg, 1994; Prevost, 1995). Variation in overall discourse structure has been found to be conveyed by intonational variation, whether in the production of DISCOURSE MARKERS or in larger patterns of variation in pitch range, pausal duration, speaking rate, and other prosodic phenomena. Finally, variation in tune or contour has been widely associated with different SPEECH ACTS in the literature. Other correlations between features such as contour, pausal duration, and final lowering with TURN-TAKING phenomena have also been studied (Sacks, Schegloff, and Jefferson, 1974; Auer, 1996; Selting, 1996; Koiso et al., 1998). And the role of intonation in conveying affect, or emotional state, is an important and still open question (Ladd et al., 1985; Cahn, 1989; Murray and Arnott, 1993; Pereira and Watson, 1998; Koike, Suzuki, and Saito, 1998; Mozziconacci, 1998).

The relation between the relative accessibility of information in a discourse and a number of observable properties of utterances has been broadly explored in theories of COMMUNICATIVE DYNAMISM, ATTENTIONAL FOCUSSING and CENTERING in discourse (Grosz, 1977; Sidner, 1979; Grosz, 1981; Sidner, 1983; Grosz, Joshi, and Weinstein, 1983; Grosz and Sidner, 1986; Chafe, 1974; Kameyama, 1986; Brennan, Friedman, and Pollard, 1987; Asher and Wada, 1988; Hajicova, Kubon, and Kubon, 1990; Gordon, Grosz, and Gillion, 1993; Gundel, Hedberg, and Zacharski, 1993), and in models of sentence production (Bock and Warren, 1985). The available evidence supports the notion that the relative accessibility of entities in the discourse model is a major factor in the assignment to grammatical role and surface position, and in the choice of the form of referring expressions: Highly accessible entities tend to be realized as the grammatical subject, to occur early in the utterance, and to be pronominalized. Furthermore, available evidence from studies on comprehension shows that accessibility is also an important factor in the way the listener processes the incoming message (Kameyama, 1986; Gordon, Grosz, and Gillion, 1993). Much research on pitch accent in discourse stems from questions of accessibility.

5.1 The Interpretation of Pronouns

While corpus-based studies have found that, on the whole, pronouns tend to be deaccented, they can be accented to convey various "marked" effects — that is, an interpretation identified in some sense as less likely. In (23), the referents of the pronouns *he* and *him* will be different in (23a) and (23b), because the accenting is different (Lakoff, 1971b).

(23) a. John called Bill a Republican and then he insulted him.

b. John called Bill a Republican and then HE insulted HIM.

In (23a), *he* and *him* are deaccented, and the likely interpretation will be that John both called Bill a Republican and subsequently insulted him. In (23b), with both pronouns accented, most hearers will understand that that John called Bill a Republican (which was tantamount to insulting him) and that Bill in return insulted John.

In another case of interaction between pitch accent and BOUND ANAPHORA, the interpretation of one clause can be affected by the intonational features of the preceding one, as in (24).

(24) a. John likes his colleagues and so does Sue.

b. John likes HIS colleagues and so does Sue.

However, this interpretation appears more clearly dependent upon the underlying semantics of the sentence and the larger context. In perception studies testing the role of accent in the STRICT/SLOPPY interpretation of ellipsis (Hirschberg and Ward, 1991a), subjects tended to favor a "marked" or less likely interpretation of sentences uttered with a pitch accent on the anaphor that they proposed for the sentence in a "neutral" (read) condition. That is, if a sentence like (24a) were likely to be interpreted with the strict reading (John likes his colleagues and Sue also likes John's colleagues), then in the spoken variant in which *his* is accented, listeners tended to favor the sloppy reading, 'John likes his own colleagues and Sue likes her own colleagues'.

Terken (1985) found that, in task-oriented monologues, speakers used deaccented, pronominal expressions to refer to the local topic of discourse, and accented, full NPs otherwise, even though many of these NPs referred to entities which had already been mentioned in the previous discourse. Pitch accent on pronouns has also been found to be correlated with changes in attentional state in studies by Cahn (1995) and by Nakatani (1997). What the conversation is "about" in terms of its topic or discourse BACKWARD-LOOKING CENTER (Grosz, Joshi, and Weinstein, 1995) can be altered, it is proposed, by the way pronouns are produced intonationally. For example, in (25), it has been suggested (Terken,

1995) that the accented pronoun in the fourth line of (25) serves to shift the topic to Betsy from Susan, who had previously been the pronominalized subject and backward-looking center of the discourse.

(25) Susan gave Betsy a pet hamster.
She reminded her such hamsters were quite shy.
She asked Betsy whether she liked the gift.
And SHE said yes, she did.
She'd always wanted a pet hamster.

5.2 The Given/New Distinction

It is a common generalization that speakers typically deaccent items that represent old, or given information in a discourse (Prince, 1981a). Mere repeated mention in a discourse is, however, clearly an inadequate definition of givenness and thus a fairly inaccurate predictor of deaccentuation. Halliday has argued that an expression may be deaccented if the information conveyed by the expression is situationally or anaphorically recoverable on the basis of the prior discourse or by being salient in the situation (Halliday, 1967). Chafe proposed that an expression may be deaccented if the information is in the listener's consciousness (Chafe, 1974; Chafe, 1976). But it seems likely that not all items which have been mentioned previously in a discourse of some length are recoverable anaphorically or are in the listener's consciousness. What is also clear is that there is no simple one-to-one mapping between givenness and deaccenting, even if givenness could be more clearly defined: Among the factors which appear to determine whether a given item is accented or not are: 1) whether or not a given item participates in a complex nominal; 2) the location of such an item in its prosodic phrase; and 3) whether preceding items in the phrase are "accentable" due to their own information status, the grammatical function of an item when first and subsequently mentioned.

For example, consider (26a) in the discourse below:

(26) a. The SENATE BREAKS for LUNCH at NOON, so i HEADED to the CAFETERIA to GET my STORY.

b. There are SENATORS, and there are THIN senators.

c. For SENATORS, LUNCH at the cafeteria is FREE. For REPORTERS, it's not.

d. But CAFETERIA food is CAFETERIA food.

(26a) shows a simple pattern of unaccented function words and accented "content words". However, in (26b), while speakers are likely to accent the content word *senators* on first mention, they are less likely to accent it on subsequent mention, when it represents given information. But in (26c), while *senators* still represents given information, speakers are likely to accent it, to contrast *senators* with *reporters*. *Cafeteria* in this utterance is likely to be deaccented, since it represents given information and is **not** being contrasted with, say, another location. But in (26d), this same given item, *cafeteria*, is likely to be accented, in part because of the stress pattern of the COMPLEX NOMINAL *cafeteria food* of which it is a part, and in part because all items in the utterance appear to be in some sense given in this context and something must bear a pitch accent in every phrase.

Brown (1983) found that all expressions used to refer to items which had been mentioned in the previous discourse were deaccented, but that expressions used to refer to inferrable items were usually accented. And Terken and Hirschberg (1994) found that differences in grammatical function of previously mentioned items with their function in the current utterance was a major factor in whether they were accented or not. It is also unclear whether what is given for a speaker, should also be treated as given for his/her illocutionary partner (Prince, 1992) — and thus, potentially deaccentable. Empirical results from the Edinburgh Map tasks dialogues (Bard, 1999) suggest that such clearly given items are rarely deaccented across speakers. And there are other studies of repeated information where it is clear that simple prior mention should not be taken as evidence of givenness for the listener, who may repeat prior data to confirm or question it (Shimojima et al., 2001).

5.3 Topic Structure

Rate, duration of inter-phrase pause, loudness, and pitch range can also convey the topic structure of a text (Silverman, 1987; Avesani and Vayra, 1988; Grosz and Hirschberg, 1992; Ayers, 1992; Swerts, Collier, and Terken, 1994; Swerts, 1997; Brown, Currie, and Kenworthy, 1980; Lehiste, 1979; Avesani and Vayra, 1988; Passoneau and Litman, 1993; Hirschberg and Nakatani, 1996; Koiso, Shimojima, and Katagiri, 1998; van Donzel, 1999). In general, it has been found that phrases beginning new topics are begun in a wider pitch range, are preceded by a longer pause, are louder, and are slower, than other phrases; narrower range, longer subsequent pause, and faster rate characterize topic-final phrases. Subsequent variation in these features then tends to be associated with a topic shift.

One of the features most frequently mentioned as important to conveying some kind of TOPIC STRUCTURE in discourse is PITCH RANGE, defined here as the distance between the maximum of the FUNDAMENTAL FREQUENCY (f0) for the vowel portions of accented syllables in the phrase, and the speaker's BASELINE, defined for each each speaker as the lowest point reached in normal speech over all. In a study of speakers reading a story, Brown et al. (1980) found that subjects typically started new topics relatively high in their pitch range and finished topics by compressing their range; they hypothesized that internal structure within a topic was similarly marked. Lehiste (1975) had reported similar results earlier for single paragraphs. Silverman (1987) found that manipulation of pitch range alone, or range in conjunction with pausal duration between utterances, could enable subjects to reliably disambiguate utterances that were intuitively potentially structurally ambiguous; for example, he used a small pitch range to signal either continuation or ending of a topic or quotation, and expanded range to indicate topic shift or quotation continuation. Avesani and Vayra (1988) also found variation in range in productions by a professional speaker which appear to correlate with topic structure, and Ayers (1992) found that pitch range appears to correlate more closely with hierarchical topic structure in read speech than in spontaneous speech. Swerts et al. (1992) also found that f0 scaling was a reliable indicator of discourse structure in spoken instructions, although the structures tested were quite simple.

Duration of pause between utterances or phrases has also been identified as an indicator of topic structure (Lehiste, 1979; Chafe, 1980; Brown, Currie, and Kenworthy, 1980; Silverman, 1987; Avesani and Vayra, 1988; Swerts, Geluykens, and Terken, 1992; Passoneau and Litman, 1993), although Woodbury (1987) found no similar correlation. Brown et al. (1980, 57) found that longer, TOPIC PAUSES (.6-.8 sec.) marked major topic shifts. Passoneau & Litman (1993) also found that the presence of a pause was a good predictor of their subjects' labeling of segment boundaries in Chafe's pear stories. Another aspect of timing, speaking rate, was found by Lehiste (1980) and by Butterworth (1975) to be associated with perception of text structure: both found that utterances beginning segments exhibited slower rates and those completing segments were uttered more rapidly.

Amplitude was also noted by Brown et al. (1980) as a signal of topic shift; they found that amplitude appeared to rise at the start of a new topic and fall at the end. Finally, contour type has been mentioned as a potential correlate of topic structure (Brown, Currie, and Kenworthy, 1980; Hirschberg and Pierrehumbert, 1986; Swerts, Geluykens, and Terken, 1992). In particular, Hirschberg and Pierrehumbert (1986) suggested that so-called DOWNSTEPPED contours[4] commonly appear either at the beginning or the ending of topics. Empirical studies showed that "low" vs. "not-low" boundary tones were good predictors of topic endings vs. continuations (Swerts, Geluykens, and Terken, 1992).

FINAL LOWERING, a compression of the pitch range during the last half second or so of an utterance, can also convey structural information to hearers, by signalling whether or not a speaker has completed his/her TURN. Pitch contour and range as well as timing have also been shown to correlate with turn-final vs. turn-keeping utterances — and distinguishing the former from discourse boundaries — as well as marking backchannels in dialogue (Sacks, Schegloff, and Jefferson, 1974; Geluykens and Swerts, 1994; Auer, 1996; Selting, 1996; Koiso et al., 1998; Caspers, 1998).

Grosz, Hirschberg, and Nakatani (1992; 1992; 1996) have also investigated the acoustic-prosodic correlates of discourse structure, inspired by the need to test potential correlates against an independent notion of discourse structure, as noted by Brown et al. (1980), and to investigate spontaneous as well as read speech. They looked at pitch range, aspects of timing and contour, and amplitude to see how well they predicted discourse segmentation decisions made by subjects using instructions based on the Grosz and Sidner (1986) model of discourse structure. They found statistically significant associations between aspects of pitch range, amplitude, and timing with segment beginnings and segment endings both for read and spontaneous speech.

Discourse Markers Accent can also disambiguate potentially ambiguous words such as DISCOURSE MARKERS, or CUE PHRASES, words and phrases such as *now*, *well*, *in the first place*. These cue phrases can function as explicit indicators of discourse structure (a discourse use) or can have a sentential reading, often as adverbials. Variation in intonational phrasing and pitch accent are correlated with the distinction between these discourse and sentential uses (Hirschberg and Litman, 1993). Tokens interpreted as discourse uses are commonly produced either as separate phrases (27a) or as part of larger phrases; in the latter case they tend to be deaccented or uttered with a **L*** accent. However, when cue phrases are produced with high prominence, they tend to be interpreted as temporal adverbs. So (27a)-(27b) are likely to be interpreted as starting a new subtopic in a discourse, while (27c) is likely to be interpreted as a temporal statement: *Now Bill is a vegetarian, although he wasn't before.* And (27a)-(27b) convey no such assertion.

(27) a. Now, Bill is a vegetarian.

b. Now Bill is a vegetarian.

c. NOW Bill is a vegetarian.

5.4 Speech Acts

There is a rich linguistic tradition characteristizing variation in overall pitch contour in many different ways: as conveying syntactic mood, speech act, speaker attitude, or speaker belief or emotion (O'Connor and Arnold, 1961; Bolinger, 1986; Bolinger, 1989; Ladd, 1980; Ladd, 1996). Some inherent meaning has often been sought in particular contours — though generally such proposals include some degree of modulation by context (Liberman and Sag, 1974; Sag and Liberman¡, 1975; Ladd, 1977; Ladd, 1978b; Bing, 1979; Ladd, 1980; Bouton, 1982; Ward and Hirschberg, 1985; Grabe et al., 1997; Gussenhoven and Rietveld, 1997). And more general attempts have been made to identify compositional meanings for contours within various systems of intonational analysis (Gussenhoven, 1983; Pierrehumbert and Hirschberg, 1990). Efforts have been made to define "standard" contours for declaratives, *wh*-questions, *yes-no* questions as a method for beginning the study of intonation in a particular language. As noted in Section 2, for example, the ToBI representation of the "standard" declarative for standard American English is **H* L-L%**, with *wh*-questions also **H* L-L%** and *yes-no* questions **L* H-H%.** Other contours' "intrinsic meaning" is, however, both more controversial and more elusive. However, below we will mention of few of the contours which have been studied by way of example.

The CONTINUATION RISE contour, which is represented by a low phrase accent and high boundary tone (**L-H%**), is generally interpreted as conveying that there is 'more to come' (Bolinger, 1989), as in (28).

(28) a. The number is L-H%: 555-1212.

b. Open the carton L-H%. Now remove the monitor carefully.

Continuation rise appears to be associated with turn-keeping phenomena, as is variation in final lowering. Internal intonational phrase boundaries in longer stretches of read speech are often realized with **L-H%**. Elements of a list, for example, are often realized as **H* L-H%** phrases.

Another contour often used in list construction is the PLATEAU contour (**H* H-L%**). However, unlike the rather neutral lists produced with continuation rise, the plateau contour conveys the sense that the speaker is talking about an 'open-ended set', as in (29).

(29) The Johnsons are solid citizens.
They **H*** pay their **H*** taxes **H-L%** .
They **H*** attend **H*** PTA meetings **H-L%** .
They're just good people.

That is, that the enumeration is for illustrative purposes only and far from complete. **H* H-L%** more generally seems to convey a certain sense that the hearer already knows the information being provided and only needs reminding — that the speaker is simply going through the motions of informing. However, this contour has received little formal study.

Much more popular among students of intonation has been the RISE-FALL-RISE contour (represented in the ToBI framework as one or more **L*+H** accents plus a low phrase accent and high boundary tone but characterized variously in other schemas), see Ladd (1980). In its more recent interpretations, it has been found to indicate either uncertainty or incredulity, depending upon the speaking rate and pitch range (See Section 2) (Ladd, 1980; Ward and Hirschberg, 1985; Hirschberg and Ward, 1992). In (30a), **L*+H L H%** is produced to indicate uncertainty; in (30b), it is produced to convey incredulity.

(30) Did you finish those slides?

a. **L*+H** Sort of **L-H%** .

b. **L*+H** Sort of **L-H%** .

Variation in aspects of pitch range and voice quality appear to be the significant factors in triggering this change in interpretation (Hirschberg and Ward, 1991b), although differences can also be observed in rate and amplitude of the two readings. "Uncertainty" interpretations have a narrower pitch range and are softer and slower than "incredulity" readings. Note that range variation can also convey differences in degree of speaker involvement, or communicate the topic structure of a text (Hirschberg and Pierrehumbert, 1986; Pierrehumbert and Hirschberg, 1990). So this type of prosodic variation can be several ways ambiguous.

L* accents can also be combined with **H*** accents to produce the so-called SURPRISE-REDUNDANCY contour (Sag and Liberman¡, 1975), as in (31); in ToBI representation, the phrase accent and boundary tone are both low.

(31) The **L*** blackboard's painted **H*** orange.

This contour has been interpreted as conveying surprise at some phenomenon that is itself observable to both speaker and hearer (hence, the notion of "redundancy"). Both this contour and a set of DOWNSTEPPED contours discussed below, might profitably be re-examined in light of currently richer resources of labeled corpora.

The downstepped contours all exhibit patterns of pitch range compression following complex pitch accents, reflected in a sequence of increasingly compressed pitch peaks in the f0 contour. In Pierrehumbert's original system, all complex pitch accents trigger downstep: **H*+L, H+L*, L*+H** and **L+H***; downstep is indicated in ToBI however by an explicit '**!**' marking on the **H** component of a downstepped pitch accent, e.g. **H* !H* !H*...**. None of the downstepped contours have been seriously studied in terms of their "meanings", although proposals have been made that **H* !H* L L%** in particular is felicitously used to open or close of a topic, especially in didactic contexts, such as academic lectures, as in (32a), or cooking classes, as in (32b).

(32) a. **H*** Today we're **!H*** going to **!H*** look at the **!H*** population of **!H*** Ghana **L-L%** .

b. **H*** This is **!H*** how you **!H*** heat the **!H*** soup **L-L%** .

In addition to investigations of contour meaning, studies have also been done on the disambiguating role various contours may play in distinguishing between DIRECT and INDIRECT speech acts — between what might be taken as the "literal meaning" of a sentence and some other illocutionary use of that sentence by a speaker (Searle, 1969). For example, a sentence with the form of a *yes-no* question, such as (33),

(33) A: Can you tell me the time?
B1: Yes.
B2: It's four o'clock.

in its literal interpretation requests a simple *yes* or *no* — is the hearer capable of providing such information? In its more customary use, however, it may be interpreted as a request to perform some action — and actually inform the questioner of the time.

Many accounts have been provided of how hearers are able to distinguish between these possible interpretations and there is considerable evidence that intonational variation can play an important role. For example, it is possible to turn a sentence with the form of a declarative into a *yes-no* question, simply by using a rising contour, as in (34):

(34) a. I like grapefruit.

b. I like grapefruit?

Most plausibly, (34a) states a fact, while (34b) seems to question a prior assertion of that fact. Perception studies performed by Sag & Liberman (1975) examined whether in fact *yes-no* questions interpreted as **direct** speech acts — requests for a simple *yes* or *no* — differed from those interpreted as **indirect** speech acts — requests to perform an action — in terms of the speaker's intonation, for sentences such as (35a). They also investigated intonational conditions under which *wh*-questions were interpreted as simple requests for information vs. those in which they were interpreted as suggestions or criticisms or denials, in sentences such as (35):

(35) a. Would you stop hitting Gwendolyn?

b. Why don't you move to California?

In preliminary findings, they reported that subjects did tend to interpret sentences like (35a) as direct speech acts when uttered with a classic interrogative contour (**L* H-H%** in ToBI notation). And productions of such sentences that were **least** likely to be interpreted as direct speech acts were uttered with a high level PLATEAU contour, e.g., (35a) uttered with the ToBI contour **H* H-L%**. *Wh*-questions such as (35b) that were interpreted as simple requests for information were often uttered with a high-low-high pattern, e.g. probably **H* L-H%** in ToBI annotation. But those interpreted as indirect speech acts — suggestions or denials — were uttered with other intonational patterns, usually falling at the end of the phrase, such as uttered as a simple declaration (**H* L-L%**). Since **H* L-L%** is thought to be the most common pattern for *wh*-questions in English, these latter findings are somewhat puzzling.

In a corpus-based study focussing on intonational features of *yes-no* questions, Steele and Hirschberg (personal communication) examined recordings of modal second-person *yes-no* questions (of the form, *Can you X?*) in recordings from a radio financial advice show. They found that tokens uttered with **L* H-H%** tended to be interpreted as requests for a simple *yes* or *no*. Tokens uttered with a standard declarative contour (**H* L-L%**) were **also** interpreted as direct speech acts — and generally answered with a simple *yes* or *no*. Utterances interpreted as indirect requests, on the other hand, tended to be those that were uttered with continuation rise (**L-H%**) or with a plateau contour (**H* H-L%**). Additionally, the modal *can* in tokens interpreted as direct was more likely to be reduced than was the modal in tokens interpreted as indirect speech acts.

Following up on this study, Nickerson and Chu-Carroll (1999) found somewhat different results in a series of production experiments. Their analysis showed that utterances realized with a low boundary tone (**L%**) were more likely to be used to convey an indirect reading (73% of tokens ending in L% were used in indirect contexts) and that those with a high boundary tone (**H%**) were slightly more likely to be used to convey a direct *yes-no* question reading (54% were used in direct contexts). So, while various studies have indeed found differences between productions of direct vs. indirect speech acts that are linked to intonational variation, the exact nature of that difference is open to further study.

Other corpus-based studies on the role of intonational variation in identifying DIALOGUE ACTS has been targeted toward speech recognition applications, but is also of some theoretical interest. Work on the DARPA Switchboard corpus (Shriberg et al., 1998) and the Edinburgh Map Task corpus (Taylor et al., 1998) has sought to associate particular intonational and lower-level prosodic features with utterances hand-labeled as, inter alia, "statements" or "acknowledgments" or BACKCHANNELS. Work on the Verbmobil corpus particularly at Erlangen (Nöth et al., 2002) has also investigated the use of prosodic features such as prominence and phrasing to improve performance in speech understanding.

6 Intonational Meaning: Future Research Areas

While there has been increasing interest in intonational studies in recent years, fueled in part by advances in the speech technologies, concern for modeling greater "naturalness": in speech synthesis (text-to-speech), and a desire to make use of whatever additional evidence intonation can provide to improve automatic speech recognition performance, much remains to be done. Corpus-based studies of all aspects of intonational meaning are still at an early stage, due to the large amount of hand labor involved in developing labeled corpora to serve as a basis for research. Study of the contribution of intonational contours to overall utterance interpretation has so far been confined to a few contours — and such common contours as continuation rise or **H* !H* L-L%** remain relatively unexamined. While there have been numerous empirical studies of accent and the given/new distinction, other forms of information status such as theme/rheme, topic/comment, and contrast could benefit from more attention. While corpus-based studies have provided some significant exceptions, most studies of intonation have examined monologue; the cross-speaker characteristics of intonation in dialogue systems offer rich prospects for investigation. And cross-language comparisons of intonational variation are also relatively scarce. In short, we still have much to learn about the pragmatics of intonation.

References

Abney, Steven. 1995. Chunks and dependencies: Bringing processing evidence to bear on syntax. In *Computational Linguistics and the Foundations of Linguistic Theory*. CSLI.

Allerton, D. J. and A. Cruttenden. 1979. Three reasons for accenting a definite subject. *Journal of Linguistics*, 15(1):49–53.

Altenberg, Bengt. 1987. *Prosodic Patterns in Spoken English: Studies in the Correlation between Prosody and Grammar for Text-to-Speech Conversion*, volume 76 of *Lund Studies in English*. Lund University Press, Lund.

Asher, N. and H. Wada. 1988. A computational account of syntactic, semantic and discourse principles for anaphora resolution. *Journal of Semantics*, 6:309–344.

Auer, Peter. 1996. On the prosody and syntax of turn-continuations. In E. Couper-Kuhlen and Margret Selting, editors, *Prosody in Conversation*. Cambridge University Press.

Avesani, Cinzia and Mario Vayra. 1988. Discorso, segmenti di discorso e un' ipotesi sull' intonazione. In *Corso di stampa negli Atti del Convegno Internazionale "Sull'Interpunzione"*, pages 8–53, Vallecchi, Firenze.

Ayers, Gayle M. 1992. Discourse functions of pitch range in spontaneous and read speech. Paper presented at the Linguistic Society of America Annual Meeting.

Baart, J. L. G. 1987. *Focus, Syntax and Accent Placement*. Ph.D. thesis, University of Leyden, Leyden.

Bachenko, Joan and Eileen Fitzpatrick. 1990. A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, 16(3):155–170.

Ball, C. N. and E. F. Prince. 1977. A note on stress and presupposition. *Linguistic Inquiry*, 8(3):585.

Bard, Ellen. 1999. The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In *Proceedings of ICPhS99*, San Francisco, August. International Congress of Phonetic Sciences.

Bardovi-Harlig, K. 1983a. Pronouns: When 'given' and 'new' coincide. In *Papers from the 18th Regional Meeting*. Chicago Linguistic Society.

Bardovi-Harlig, Kathleen. 1983b. *A Functional Approach to English Sentence Stress*. Ph.D. thesis, University of Chicago, Chicago IL.

Beach, Cheryl. 1991. The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30:644–663.

Bing, Janet. 1979. *Aspects of English Prosody*. Ph.D. thesis, University of Massachusetts at Amherst, Amherst MA. Distributed by the Indiana University Linguistics Club.

Birch, S. and C. Clifton. 1995. Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, 38(4):365–391.

Bock, J. K. and R. K. Warren. 1985. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21:47–67.

Bolinger, Dwight. 1972. Accent is predictable (if you're a mindreader). *Language*, 48:633–644.

Bolinger, Dwight. 1986. *Intonation and Its Parts: Melody in Spoken English.* Stanford University Press, Palo Alto CA.

Bolinger, Dwight. 1989. *Intonation and Its Uses: Melody in Grammar and Discourse.* Edward Arnold, London.

Boula de Mareüil, Philippe and Christophe d'Alessandro. 1998. Text chunking for prosodic phrasing in french. In *The Third ESCA/COCOSDA Workshop on Speech Synthesis*, pages 127–131, Jenolan Caves Mountain House, Blue Mountains, Australia, November.

Bouton, Lawrence F. 1982. Stem polarity and tag intonation in the derivation of the imperative tag. In Robinson Schneider, Kevin Tuite, and Robert Chametzky, editors, *Papers from the Parasession on Nondeclaratives*, pages 23–42, Chicago. Chicago Linguistic Society.

Brennan, Susan E., Marilyn W. Friedman, and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting*, pages 155–162, Stanford. Association for Computational Linguistics.

Bresnan, Joan. 1971. Sentence stress and syntactic transformations. *Language*, 47:257–281.

Brown, G. 1983. Prosodic structure and the given/new distinction. In D. R. Ladd and A. Cutler, editors, *Prosody: Models and Measurements.* Springer Verlag, Berlin, pages 67–78.

Brown, G., K. Currie, and J. Kenworthy. 1980. *Questions of Intonation.* University Park Press, Baltimore.

Büring, Daniel. 1999. Topic. In *Focus – Linguistic, Cognitive, and Computational Perspectives.* Cambridge University Press.

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

Cahn, Janet. 1998. Generating pitch accent distributions that show individual and stylistic differences. In *The Third ESCA/COCOSDA Workshop on Speech Synthesis*, pages 121–126, Jenolan Caves Mountain House, Blue Mountains, Australia, November.

Cahn, Janet E. 1989. Generating expression in synthesized speech. Master's thesis, Massachusetts Institute of Technology, May.

Cahn, Janet E. 1995. The effect of pitch accenting on pronoun referent resolution. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (Student Session)*, pages 290–292.

Campbell, Nick and Mary Beckman. 1997. Stress, prominence, and spectral tilt. In A. Botinis, G. Kouroupetroglou, and G. Carayiannis, editors, *Intonation: Theory, Models and Applications*, pages 67–70, Athens. ESCA.

Caspers, Johanneke. 1998. Who's next? the melodic marking of question vs. continuation in dutch. *Language and Speech: Special Issue on Prosody and Conversation*, 41(3-4).

Chafe, W. L. 1980. The deployment of consciousness in the production of a narrative. In W. L. Chafe, editor, *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Navvative Production*. Ablex Publishing Corp, pages 9–50. Vol. 3, Advances in Discourse Processes.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li, editor, *Subject and Topic*. The Academic Press, New York, pages 25–55.

Chafe, Wallace L. 1974. Language and consciousness. *Language*, 50:111–133.

Cooper, William E. and Jeanne Paccia-Cooper. 1980. *Syntax and Speech*. Harvard University Press, Cambridge MA.

Croft, William. 1995. Intonation units and grammatical structure. *Linguistics*, 33:839–882.

Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge University Press, Cambridge UK.

Culicover, Peter W. and Michael Rochemont. 1983. Stress and focus in English. *Language*, 59(1):123–165, March.

Cutler, A. and D. Foss. 1977. On the role of sentence stress in sentence processing. *Language and Speech*, 20:1–10.

Dirksen, A. 1992. Accenting and deaccenting: A declarative approach. In *Proceedings of COLING-92*, pages 865–869.

Dirksen, Arthur and Hugo Quene. 1993. Prosodic analysis: The next generation. In Vincent J. van Hueven and Louis C. W. Pols, editors, *Analysis and Synthesis of Speech: Strategic Research towards High-Quality Text-to-Speech Generation*. Mouton de Gruyter, pages 131–144.

Downing, Bruce T. 1970. *Syntactic Structure and Phonological Phrasing in English*. Ph.D. thesis, University of Texas, Austin.

Eady, S. J. and W. E. Cooper. 1986. Speech intonation and focus location in matched statements & questions. *Journal of the Acoustical Society of America*, 80:402–415.

Enkvist, N. 1979. Marked focus: Functions and constraints. In S. Greenbaum, G. Leech, and J. Svartvik, editors, *Studies in English Linguistics for Randolph Quirk*. Longmans, London, pages 134–152.

Erteschik-Shir, Nomi and Shalom Lappin. 1983. Under stress: A functional explanation of English sentence stress. *Journal of Linguistics*, 19:419–453.

Fowler, C. A. and J. Housum. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, 26:489–504.

Fuchs, A. 1980. Accented subjects in 'all-new' utterances. In G. Brettschneider and C. Lehmann, editors, *Wege zur Universalienforschung: sprachwissenschaftliche Beitrage zum 60.* Narr, Tubingen.

Fuchs, A. 1984. Deaccenting and default accent. In D. Gibbon and H. Richter, editors, *Intonation, Accent and Rhythm.* Walter de Gruyter, Berlin, pages 134–164.

Fujio, Shigeru, Yoshinori Sagisaka, and Norio Higuchi. 1997. Prediction of major phrase boundary location and pause insertion using a stochastic context-free grammar. In Yoshinori Sagisaka, Nick Campbell, and Norio Higuchi, editors, *Computing Prosody: Computational Models for Processing Spontaneous Speech*. Springer, pages 271–283.

Gee, J. P. and F. Grosjean. 1983. Performance structure: A psycholinguistic and linguistic apprasial. *Cognitive Psychology*, 15:411–458.

Geluykens, Ronald and Marc Swerts. 1994. Prosodic cues to discourse boundaries in experimental dialogues. *Speech Communication*, 15:69–77.

Gordon, P. C., B. J. Grosz, and L. A. Gillion. 1993. Pronouns, names, and the centering of attention in discourse. *Journal of Psycholinguistic Research*.

Grabe, E., C. Gussenhoven, J. Haan, E. Marsi, and B. Post. 1997. The meaning of intonation phrase onsets in Dutch. In A. Botinis, G. Kouroupetroglou, and G. Carayiannis, editors, *Intonation: Theory, Models and Applications*, pages 161–164, Athens. ESCA.

Grosjean, F., L. Grosjean, and H. Lane. 1979. The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11:58–81.

Grosz, B., A.K. Joshi, and S. Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting*, pages 44–50, Cambridge MA, June. Association for Computational Linguistics.

Grosz, B. J. 1981. Focusing and description in natural language dialogues. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*. Cambridge University Press, Cambridge, England.

Grosz, B. J., A. K. Joshi, and S. Weinstein. 1995. Centering: a framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–25.

Grosz, Barbara and Julia Hirschberg. 1992. Some intonational characteristics of discourse structure. In *Proceedings of ICSLP-92*, Banff, October.

Grosz, Barbara J. 1977. The representation and use of focus in dialogue understanding. Technical Report 151, SRI International, Menlo Park Ca. University of California at Berkeley PhD Thesis.

Grosz, Barbara J. and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Gundel, J. 1978. Stress, pronominalization, and the given-new distinction. *University of Hawaii Working Papers in Linguistics*, 10(2):1–13.

Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):272–307.

Gussenhoven, Carlos. 1983. *On the Grammar and Semantics of Sentence Accents*. Foris Publications, Dordrecht.

Gussenhoven, Carlos and Toni Rietveld. 1997. Empirical evidence for the contrast between L* and H* in dutch rising contours. In A. Botinis, G. Kouroupetroglou, and G. Carayiannis, editors, *Intonation: Theory, Models and Applications*, pages 18–20, Athens. ESCA.

Hajicova, E., P. Kubon, and V. Kubon. 1990. Hierarchy of salience and discourse analysis and production. In *Papers Presented to the 13th International Conference on Computational Linguistics*, pages 144–148, Helsinki. International Conference on Computational Linguistics.

Halliday, M. A. K. 1967. Notes on transitivity and theme in English, part 2. *Journal of Linguistics*, 3:199–244.

Hirschberg, J. and C. Avesani. 2000. Prosodic disambiguation in English and Italian. In A. Botinis, editor, *Intonation: Analysis, Modelling and Technology*. Kluwer.

Hirschberg, J. and Gregory Ward. 1992. The influence of pitch range, duration, amplitude, and spectral features on the interpretation of **l*+h l h%**. *Journal of Phonetics*, 20(2):241–251.

Hirschberg, Julia. 1993. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, 63:305–340, October.

Hirschberg, Julia and Cinzia Avesani. 1997. The role of prosody in disambiguating potentially ambiguous utterances in English and Italian. In *ESCA Tutorial and Research Workshop on Intonation: Theory, Models and Applications*, Athens, September.

Hirschberg, Julia and Barbara Grosz. 1992. Intonational features of local and global discourse structure. In *Proceedings of the Speech and Natural Language Workshop*, pages 441–446, Harriman NY, February. DARPA, Morgan Kaufmann.

Hirschberg, Julia and Diane Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3):501–530.

Hirschberg, Julia and Christine Nakatani. 1996. A prosodic analysis of discourse segments in direction-giving monologues. In *Proceedings of the 34th Annual Meeting*, Santa Cruz. Association for Computational Linguistics.

Hirschberg, Julia and Janet Pierrehumbert. 1986. The intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting*, pages 136–144, New York. Association for Computational Linguistics.

Hirschberg, Julia and Pilar Prieto. 1996. Training intonational phrasing rules automatically for English and Spanish text-to-speech. *Speech Communication*, 18:281–290.

Hirschberg, Julia and Gregory Ward. 1991a. Accent and bound anaphora. *Cognitive Linguistics*, 2(2):101–121.

Hirschberg, Julia and Gregory Ward. 1991b. The influence of pitch range, duration, amplitude, and spectral features on the interpretation of **l*+h l h%**. *Journal of Phonetics*.

Horne, M. 1985. English sentence stress, grammatical functions and contextual coreference. *Studia Linguistica*, 39:51–66.

Horne, M. 1991a. Accentual patterning in 'new' vs 'given' subjects in English. Working Papers 36, Department of Linguistics, Lund University, Lund.

Horne, M. 1991b. Phonetic correlates of the new/given parameter. In *Proceedings of the Twelfth International Congress of Phonetic Sciences*, pages 230–233, Aix-en-Provence. ICPhS.

Horne, Merle. 1987. Towards a discourse-based model of English sentence intonation. Working Papers 32, Lund University Department of Linguistics.

Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge MA.

Kameyama, Megumi. 1986. A property-sharing constraint in centering. In *Proceedings of ACL-86*, pages 200–206, Cambridge, MA.

Koike, Kazuhito, Hirotaka Suzuki, and Hiroaki Saito. 1998. Prosodic parameters in emotional speech. In *Proceedings of ICSLP-98*, Sydney.

Koiso, Hanae, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa, and Yasuharu Den. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map task dialogs. *Language and Speech: Special Issue on Prosody and Conversation*, 41(3-4).

Koiso, Hanae, Atsushi Shimojima, and Yasuhiro Katagiri. 1998. Collaborative signaling of informational structures by dynamic speech rate. *Language and Speech: Special Issue on Prosody and Conversation*, 41(3-4).

Kruyt, J. Geertruida. 1985. *Accents from Speakers to Listeners: An Experimental Study of the Production and Perception of Accent Patterns in Dutch*. Ph.D. thesis, University of Leyden.

Kutik, E. J., W. E. Cooper, and S. Boyce. 1983. Declination of fundamental frequency in speaker's production of parenthetical and main clauses. *Journal of the Acoustical Society of America*, 73:1731–1738.

Ladd, D. R. 1979b. Light and shadow: A study of the syntax and semantics of sentence accents in English. In L. Waugh and F. van Coetsem, editors, *Contributions to Grammatical Studies: Semantics and Syntax*. University Park Press, Baltimore, pages 93–131.

Ladd, D. R., K. E. A. Silverman, F. Tolkmitt, G. Bergmann, and K. R. 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(1):435–444, August.

Ladd, D. Robert. 1977. The function of the a-rise accent in English. Distributed by the Indiana University Linguistics Club.

Ladd, D. Robert. 1978. Stylized intonation. *Language*, 54:517–540.

Ladd, D. Robert. 1980. *The Structure of Intonational Meaning*. Indiana University Press, Bloomington, Ind.

Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge University Press.

Lakoff, George. 1971. Presupposition and relative well-formedness. In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. Cambridge University Press, Cambridge UK, pages 329–340.

Lehiste, I. 1975. The phonetic structure of paragraphs. In A. Cohen and S. G. Nooteboom, editors, *Structure and Process in Speech Perception*. Springer, Heidelberg, pages 195–203.

Lehiste, I. 1979. Perception of sentence and paragraph boundaries. In B. Lindblom and S. Oehman, editors, *Frontiers of Speech Research*. Academic Press, London, pages 191–201.

Lehiste, I. 1980. Phonetic characteristics of discourse. Paper presented at the Meeting of the Committee on Speech Research, Acoustical Society of Japan.

Lehman, Christina. 1977. A re-analysis of givenness: Stress in discourse. In *Papers from the 13th Annual Meeting*, pages 316–24. Chicago Linguistic Society.

Liberman, M. and R. Sproat. 1992. The stress and structure of modified noun phrases in English. In I. Sag, editor, *Lexical Matters*. University of Chicago Press, pages 131–182.

Liberman, Mark and Ivan A. Sag. 1974. Prosodic form and discourse function. In *Papers of the Tenth Regional Meeting*, pages 416–427. Chicago Linguistic Society.

Marcus, Mitchell P. and Donald Hindle. 1990. Description theory and intonation boundaries. In Gerald Altmann, editor, *Computational and Cognitive Models of Speech*. MIT Press, Cambridge MA.

Mozziconacci, S. J. L. 1998. *Speech Variability and Emotion: Production and perception*. Ph.D. thesis, University of Leiden.

Murray, Iain R. and Mohn L. Arnott. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93(2):1097–1108.

Nakatani, Christine. 1997. *The Computational Processing of Intonational Prominence: A Functional Prosody Perspective*. Ph.D. thesis, Harvard University.

Nespor, M. and I. Vogel. 1983. Prosodic structure above the word. In A. Cutler and D. Robert Ladd, editors, *Prosody: Models and Measurements*. Springer-Verlag, pages 123–132.

Nickerson, J. S. and J. Chu-Carroll. 1999. Acoustic-prosodic disambiguation of direct and indirect speech acts. In *Proceedings of the XIV International Congress of Phonetic Sciences*, pages 1309–12, San Francisco.

Nooteboom, S. G. and J. Terken. 1982. What makes speakers omit pitch accents?: An experiment. *Phonetica*, 39:317–336.

Nöth, Elmar, Anton Batliner, V. Warnke, J. Haas, M. Boros, J. Buckow, R. Huber, F. Gallwitz, M. Nutt, and H. Niemann. 2002. On the use of prosody in automatic dialogue understanding. *Speech Communication*, 36.

O'Connor, J.D. and G.F. Arnold. 1961. *Intonation of Colloquial English*. Longman, London.

Oehrle, Richard T. 1991. Prosodic constraints on dynamic grammar analysis. In S. Bird, editor, *Declarative Perspectives on Phonology*, volume 7. Centre for Cognitive Science, University of Edinburgh, pages 167–95. Working Papers in Cognitive Science.

Ostendorf, M. and N. Veilleux. 1994. A hierarchical stochastic model for automatic prediction of prosodic boundary location. *Computational Linguistics*, 20(1):27–54.

Partee, Barbara. 1991. Topic focus and quantification. In Steven Moore and Adam Z. Wyner, editors, *Proceedings from the First Conference on Semantic and Linguistic Theory: SALT*. Cornell Working Papers in Linguistics.

Passoneau, R. and D. Litman. 1993. Feasibility of automated discourse segmentation. In *Proceedings of the 31st Annual Meeting*, Ohio State University. Association for Computational Linguistics.

Pereira, Cecile and Catherine Watson. 1998. Some acoustic characteristics of emotion. In *Proceedings of ICSLP-98*, Sydney.

Pierrehumbert, Janet and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge MA, pages 271–311.

Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. thesis, Massachusetts Institute of Technology, September. Distributed by the Indiana University Linguistics Club.

Pitrelli, John, Mary Beckman, and Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of ICSLP-94*, volume 2, pages 123–126, Yokohama.

Prevost, Scott. 1995. *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation*. Ph.D. thesis, University of Philadelphia.

Prevost, Scott and Mark Steedman. 1994. Specifying intonation from context for speech synthesis. *Speech Communication*, 15:139–153.

Price, P. J., M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. 1990. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, December.

Prince, E. F. 1992. The ZPG letter: Subjects, definiteness, and information-status. In S. Thompson and W. Mann, editors, *Discourse Description: Diverse Analyses of a Fund Raising Text*. John Benjamins B. V., Philadelphia, pages 295–325.

Prince, E.F. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*. The Academic Press, New York, pages 223–255.

Rochemont, Michael S. and Peter W. Culicover. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge University Press, Cambridge UK.

Rooth, Mats. 1985. *Association with Focus.* Ph.D. thesis, University of Massachusetts, Amherst MA.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1:75–116.

Sacks, H., E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking in conversation. *Language*, 50(4):696–735.

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

Schafer, A. J., S. R. Speer, P. Warren, and S. D. White. 2000. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, 29:169–182.

Schmerling, S. 1971. Presupposition and the notion of normal stress. In *Papers from the 7th Regional Meeting*, pages 242–253. Chicago Linguistic Society.

Schmerling, S. 1974. A re-examination of the notion NORMAL STRESS. *Language*, 50:66–73.

Schmerling, S. 1975. Evidence from sentence stress for the notions of topic and comment. In S. Schmerling and R. King, editors, *Texas Linguistics Forum.* University of Texas, Department of Linguistics, pages 135–141.

Schmerling, Susan F. 1976. *Aspects of English Sentence Stress.* University of Texas Press, Austin. Revised 1973 thesis, University of Illinois at Urbana.

Schwarzschild, Roger. 1999. GIVENness, avoidf and other constraints on the placement of focus. *Natural Language Semantics.*

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge UK.

Selkirk, E. 1984. *Phonology and Syntax: The Relation between Sound and Structure.* MIT Press, Cambridge MA.

Selkirk, E. 1995. Sentence prosody: Intonation, stress and phrasing. In John Goldsmith, editor, *Handbook of Phonological Theory.* Basil Blackwell, pages 550–69.

Selting, Margret. 1996. On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics*, 6(3):371–389.

Sgall, Petr, Eva Hajičova, and Jarmila Panevová. 1986. *The Meaning of the Sentence in the Semantic and Pragmatic Aspects.* Academia, Prague.

Shimojima, Atsushi, Yasuhiro Katagiri, Hanae Koiso, and Marc Swerts. 2001. An experimental study on the informational and grounding functions of prosodic features of Japanese echoic responses. *Speech Communication*. An extended version of a paper present at the ESCA Workshop on Dialogue and Prosody, Veldhoven, 1999.

Shriberg, Elizabeth, Rebecca Bates, Andreas Stolcke, Paul Taylor, Daniel Jurafsky, Klaus Ries, Noah Coccaro, Rachel Martin, Marie Meteer, and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech: Special Issue on Prosody and Conversation*, 41(3-4):443–492.

Sidner, C. 1979. *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse*. Ph.D. thesis, The Massachusetts Institute of Technology.

Sidner, C. L. 1983. Focusing in the comprehension of definite anaphora. In M. Brady and R. Berwick, editors, *Computational Models of Discourse*. MIT Press, Cambridge MA, pages 267–330.

Silverman, K. 1987. *The Structure and Processing of Fundamental Frequency Contours*. Ph.D. thesis, Cambridge University, Cambridge UK.

Silverman, K., M. Beckman, J. Pierrehumbert, M. Ostendorf, C. Wightman, P. Price, and J. Hirschberg. 1992. ToBI: A standard scheme for labeling prosody. In *Proceedings of ICSLP-92*, pages 867–879, Banff, October.

Sproat, Richard. 1994. English noun-phrase accent prediction for text-to-speech. *Computer Speech and Language*, 8:79–94.

Steedman, Mark. 1991. Structure and intonation. *Language*, 67:260–296.

Swerts, M. 1997. Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, 22(1):25–41.

Swerts, M., R. Geluykens, and J. Terken. 1992. Prosodic correlates of discourse units in spontaneous speech. In *Proceedings*, pages 421–428, Banff, October. International Conference on Spoken Language Processing.

Swerts, Marc, Rene Collier, and Jacques Terken. 1994. Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication*, 15:79–90.

Taylor, Paul, Simon King, Stephen Isard, and Helen Wright. 1998. Intonation and dialog context as constraints for speech recognition. *Language and Speech: Special Issue on Prosody and Conversation*, 41(3-4).

Terken, J. 1984. The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, 27:269–289.

Terken, J. and S. G. Nooteboom. 1987. Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, 2(3/4):145–163.

Terken, J. M. B. 1985. *Use and Function of Accentuation: Some Experiments.* Ph.D. thesis, University of Leiden, Helmond, Neth.

Terken, Jacques. 1995. Accessibility, prominence, pronouns and accents. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering in Discourse.* Oxford University Press.

Terken, Jacques. 1997. Variation of accent prominence within the phrase: Models and spontaneous speech data. In Yoshinori Sagisaka, Nick Campbell, and Norio Higuchi, editors, *Computing Prosody: Computational Models for Processing Spontaneous Speech.* Springer, pages 95–116.

Terken, Jacques and Julia Hirschberg. 1994. Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position. *Language and Speech*, 37(2):125–145.

Vallduvì, Enric. 1999. *The Informational Component.* Garland, New York.

van Donzel, Monique. 1999. *Prosodic Aspects of Information Structure in Discourse.* Holland Academic Graphics.

Veilleux, Nanette. 1994. *Computational Models of the Prosody/Syntax Mapping for Spoken Language Systems.* Ph.D. thesis, Boston University.

Wales, Roger and Hugh Toner. 1979. Intonation and ambiguity. In W. E. Cooper and E. C. Walker, editors, *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett.* Halsted Press, New York, pages 135–58.

Ward, G. and J. Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, 61(4):747–776, December.

Wilson, Dierdre and Dan Sperber. 1979. Ordered entailments: An alternative to presuppositional theories. In C.-K. Oh and D. A. Dinneen, editors, *Syntax and Semantics*, volume 11. The Academic Press, New York, pages 229–324.

Woodbury, Anthony C. 1987. Rhetorical structure in a central Alaskan Yupik Eskimo traditional narrative. In J. Sherzer and A. Woodbury, editors, *Native American Discourse: Poetics and Rhetoric.* Cambridge University Press, Cambridge UK, pages 176–239.

Zacharski, Ron. 1992. Generation of accent in nominally premodified noun phrases. In *Papers Presented to the 15th International Conference on Computational Linguistics*, pages 253–259, Nantes. International Conference on Computational Linguistics.

Notes

[1]A fuller description of the ToBI systems may be found in the ToBI conventions document and the training materials available at http://ling.ohio-state.edu/ tobi. Other versions of this system have been developed for languages such as German, Italian, Japanese, and Spanish.

[2]The examples in Figures 1–5 are taken from the ToBI training materials, prepared by Mary Beckman and Gail Ayers, and available at http://ling.ohio-state.edu/ tobi.

[3]Pierrehumbert's **H+L*** corresponds to the ToBI +̋!H*. Her **H*+L** is included in the simple **H*** category, and may be distinguished contextually from the simple **H*** by the presence of a following down-stepped tone. Otherwise the systems are identical.

[4]Contours in which one or more pitch accents which follow a complex accent are uttered in a compressed range, producing a "stairstep" effect.