Empirical Studies on the Disambiguation of Cue Phrases

Julia Hirschberg
Diane Litman
AT&T Bell Laboratories
600 Mountain Avenue Murray Hill, NJ 07974
{julia,diane}@research.att.com

June 21, 1995

Abstract

Cue phrases are linguistic expressions such as *now* and *well* that function as explicit indicators of the structure of a discourse. For example, *now* may signal the beginning of a subtopic or a return to a previous topic, while *well* may mark subsequent material as a response to prior material, or as an explanatory comment. However, while cue phrases may convey discourse structure, each also has one or more alternate uses. While *incidentally* may be used **sententially** as an adverbial, for example, the **discourse** use initiates a digression. Although distinguishing discourse and sentential uses of cue phrases is critical to the interpretation and generation of discourse, the question of how speakers and hearers accomplish this disambiguation is rarely addressed.

This paper reports results of empirical studies on discourse and sentential uses of cue phrases, in which both text-based and prosodic features were examined for disambiguating power. Based on these studies, it is proposed that discourse versus sentential usage may be distinguished by intonational features, specifically, **pitch accent** and **prosodic phrasing**. A prosodic model is identified that characterizes these distinctions. This model is associated with features identifiable from text analysis, including orthography and part-of-speech, to permit the application of the results of the prosodic analysis to the generation of appropriate intonational features for discourse and sentential uses of cue phrases in synthetic speech.

1 Introduction

Cue phrases, words and phrases that directly signal the structure of a discourse, have been variously termed **clue words**, **discourse markers**, **discourse connectives**, and **discourse particles** in the computational linguistic and conversational analysis literature. These include items such as *now*, which marks the introduction of a new subtopic or return to a previous one; *well*, which indicates a response to previous material or an explanatory comment; *incidentally*, *by the way*, and *that reminds me*, which indicate the beginning of a digression; and *anyway* and *in any case*, which indicate a return from a digression. The recognition and appropriate generation of cue phrases is of particular interest to research in discourse structure. The structural information conveyed by these phrases is crucial to many tasks, such as anaphora resolution (Grosz, 1977; Grosz and Sidner, 1986; Reichman, 1985), the inference of speaker intention and the recognition of speaker plans (Grosz and Sidner, 1986; Sidner, 1985; Litman and Allen, 1987), and the generation of explanations and other text (Zuckerman and Pearl, 1986).

Despite the crucial role that cue phrases play in theories of discourse and their implementation, however, many questions about how cue phrases are identified and defined remain to be examined. In particular, the question of cue phrase polysemy has yet to receive a satisfactory solution. Each lexical item that has one or more **discourse** senses also has one or more alternate, **sentential** senses, which make a semantic contribution to the interpretation of an utterance. So, sententially, *now* may be used as a temporal adverbial, *incidentally* may also function as an adverbial, and *well* may be used with its adverbial or attributive meanings. Distinguishing between whether a discourse or a sentential usage is meant, is, obviously, critical to the interpretation of discourse.

Consider the cue phrase *now*. Roughly, the sentential or deictic use of *now* makes reference to a span of time which minimally includes the utterance time. This time span may include little more than moment of utterance, as in Example (1), or it may be of indeterminate length, as in Example (2).

(1) **Fred**: Yeah I think we'll look that up and possibly uh after one of your breaks Harry.
Harry: OK we'll take one *now*. Just hang on Bill and we'll be right back with you.

(2) **Harry:** You know I see more coupons *now* than I've ever seen before and I'll bet you have too.

These examples are taken from a radio call-in program, "The Harry Gross Show: Speaking of Your Money" (Pollack, Hirschberg, and Webber, 1982), which we will refer to as (HG82). This corpus will be described in more detail in Section 4.

In contrast, the discourse use of *now* signals a return to a previous topic, as in the two examples of *now* in Example (3) (HG82), or introduces a subtopic, as in Example (4) (HG82).

(3) **Harry:** Fred whatta you have to say about this IRA problem?
Fred: Ok. You see *now* unfortunately Harry as we alluded to earlier when there is a distribution from an IRA that is taxable ...discussion of caller's beneficiary status... *Now* the the five thousand that you're alluding to uh of the —

(4) **Doris:** I have a couple quick questions about the income tax. The first one is my husband is retired and on social security and in '81 he . . . few odd jobs for a friend uh around the property and uh he was reimbursed for that to the tune of about 640 dollars. *Now* where would he where would we put that on the form?

Example (5) nicely illustrates both the discourse and sentential uses of *now* in a single utterance.

(5) *Now now* that we have all been welcomed here it's time to get on with the business of the conference.

In particular, the first *now* illustrates a discourse usage, and the second a sentential usage. This example is taken from a keynote address given by Ronald Brachman to the *First International Conference on Expert Database Systems* in 1986. We will refer to this corpus as (RJB86). The corpus will be described in more detail in Section 5.

While the distinction between discourse and sentential usages sometimes seems quite clear from context, in many cases it is not. From the text alone, Example (6) (RJB86) is potentially ambiguous between a temporal reading of *now* and a discourse interpretation.

(6) *Now* in AI our approach is to look at a knowledge base as a set of symbolic items that represent something.

On the temporal reading, Example (6) would convey that 'at this moment the AI approach to knowledge bases has changed'; on the discourse reading, *now* simply initiates the topic of 'the AI approach to knowledge bases'.

In this paper, we address the problem of disambiguating cue phrases in both text and speech. We present results of several studies of cue phrase usage in corpora of recorded, transcribed speech, in which we examined text-based and prosodic features to find which best predicted the discourse/sentential distinction. Based on these analyses, we present an intonational model for cue phrase disambiguation in speech, based on prosodic phrasing and pitch accent. We associate this model with features identifiable from text analysis, principally, orthography and part-of-speech, that can be automatically extracted from large corpora. On a practical level, this association permits the application of our findings to the identification and appropriate generation of cue phrases in synthetic speech. On a more theoretical level, our findings provide support for theories of discourse which rely upon the feasibility of cue phrase disambiguation to support the identification of discourse structure. Our results provide empirical evidence suggesting how hearers and readers may distinguish between discourse and sentential uses of cue phrases. More generally, our findings can be seen as a case study demonstrating the importance of intonational information to language understanding and generation.

In Section 2 we review previous work on cue phrases and discuss the general problem of distinguishing between discourse and sentential uses. In Section 3 we introduce the theory of English intonation adopted for our prosodic analysis (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986). In Section 4 we present our initial empirical studies, which focus on the analysis of the cue phrases *now* and *well* in multi-speaker spontaneous speech. In Section 5 we demonstrate that these results generalize to other cue phrases, presenting results of a larger and more comprehensive study: an examination of all cue phrases produced by a single speaker in a seventy-five minute presentation. Finally, in Section 6 we discuss the theoretical and practical applications of our findings.

2 Previous Studies of Cue Phrases

The critical role that cue phrases play in understanding and generating discourse has often been noted in the computational linguistics literature. For example, it has been shown that cue phrases can assist in the resolution of anaphora, by indicating the presence of a structural boundary or a relationship between parts of a discourse (Grosz, 1977; Grosz and Sidner, 1986; Reichman, 1985). In Example (7) (RJB86), interpretation of the anaphor *it* as co-indexed with *the system* is facilitated by the presence of the cue phrases *say* and *then*, marking potential antecedents in "as *an expert database* for *an expert system*" as structurally unavailable.

(7) If *the system* attempts to hold rules, *say* as *an expert database* for *an expert system*, *then* we expect *it* not only to hold the rules but to in fact apply them for us in appropriate situations.

Here, *say* indicates the beginning of a discourse subtopic and *then* signals a return from that subtopic. Since the potential but incorrect antecedents occur in the subtopic, while the pronoun in question appears in the return to the major topic, the incorrect potential antecedents can be ruled out on structural grounds. Without such discourse segmentation, the incorrect potential antecedents might have been preferred, given their surface proximity and number agreement with the pronoun in question. Note that, without cue phrases as explicit indicators of this topic structure, one would have to infer the relationships among discourse segments by appeal to a more detailed analysis of the semantic content of the passage. For example, in task-oriented dialogues, plan-based knowledge could be used to assist in the recognition of discourse structure (Grosz, 1977). However, such analysis is often beyond the capabilities of current natural language processing systems. Many domains are also not task-oriented. Additionally, cue phrases are widely used in the identification of rhetorical relations among portions of a text or discourse (Hobbs, 1979; Mann and Thompson, 1983; Reichman, 1985), and have been claimed in general to reduce the complexity of discourse processing and to increase textual coherence in natural language processing systems (Cohen, 1984; Litman and Allen, 1987; Zuckerman and Pearl, 1986).

Previous attempts to characterize the set of cue phrases in the linguistic and in the computational literature have typically been extensional, with each cue phrase or set of phrases associated with one or more discourse or conversational functions. In the linguistic literature, cue phrases have been the subject of a number of theoretical and descriptive corpus-based studies, which emphasize the diversity of meanings associated with cue phrases as a class, within an overarching framework of function such as **discourse cohesiveness** or **conversational moves**, and the diversity of meanings that an individual item can convey (Halliday and Hassan, 1976; Schiffrin, 1987; Schourup, 1985; Warner, 1985).

In the computational literature, the functions assigned to each cue phrase, while often more specific than those identified in the linguistics literature, are usually theory or domain dependent. Reichman (1985) and Hobbs (1979) associate groups of cue phrases with the rhetorical relations among segments of text which they signal; in these approaches, the cue phrase taxonomy is dependent upon the set of rhetorical relations assumed. Alternatively, Cohen (1984) adopts a taxonomy of connectives based on (Quirk, 1972) to assign each class of cue phrase a function in her model of argument understanding. Grosz and Sidner (1986), in their tripartite model of discourse structure, classify cue phrases based on the changes they signal to the attentional and intentional states. Zukerman (1986) presents a taxonomy of cue phrases based on three functions in the generation of tutorial explanations: knowledge organization, knowledge acquisition, and affect maintenance. Table 14 in the Appendix compares the characterization of items classed as cue phrases in a number of these classification schemes.

The question of cue phrase sense ambiguity has been noted in both the computational and the linguistic literature, although only cursory attention has been paid to how disambiguation might take place. A common assumption in the computational literature is that hearers can use surface position within a sentence or clause to distinguish discourse from sentential uses. In fact, most systems that recognize or generate cue phrases assume a canonical (usually first) position for discourse cue phrases within the clause (Reichman, 1985; Zuckerman and Pearl, 1986). Schiffrin (1987) also assumes that discourse uses of cue phrases are utterance initial.

However, discourse uses of cue phrases can in fact appear non-initially in a clause, as illustrated by the item *say* in Example (8) (RJB86).

(8) However, if we took that language and added one simple operator which we called restriction which allowed us for example to form relational concepts like *say*, son and daughter, that is a child who is always male or is always female.

Also, sentential usages can appear clause initially, as in Example (9) (RJB86).

(9) We've got to get to some inferential capability. *Further* meaning of the structures is crucially important.

Furthermore, surface clausal position itself may be ambiguous in the absence of orthographic disambiguation. Consider Example (10) (HG82).

(10) **Evelyn**: I see. So in other words I will have to pay the full amount of the uh of the tax *now* what about Pennsylvania state tax? Can you give me any information on that?

Here, *now* would be assigned a sentential interpretation if associated with the preceding clause, *I will have to pay the full amount of the...tax now*, but a discourse interpretation if associated with the succeeding clause, *Now what about Pennsylvania state tax?* Thus, surface position alone appears inadequate to distinguish between discourse and sentential usage.

However, when we listen to examples such as Example (10), we have little difficulty in identifying a discourse meaning for *now*. Similarly, the potentially troublesome case cited in Example (6) is easily disambiguated when one listens to the recording itself. What is missing from transcription that helps listeners to make such distinctions easily?

Halliday and Hassan (1976, p.268) note that their class of **continuatives**, which includes items such as *now*, *of course*, *well*, *anyway*, *surely*, and *after all* (i.e. items also commonly classed as cue phrases), vary intonationally with respect to cohesive function. In particular, continuatives are often "reduced" intonationally when they function "cohesively" to relate one part of a text to another (i.e. in their discourse use), unless they are "very definitely contrastive"; that is, continuatives are unaccented, with reduced vowel forms, unless they are marked as unusually prominent intonationally. For example, they note that, if *now* is reduced, it can indicate "the opening of a new stage in the communication", such as a new point in an argument or a new incident in a story. On the other hand, non-cohesive uses, which we would characterize as sentential, tend to be of non-reduced, accented forms.

So, perhaps it is the intonational information present in speech, but missing generally in transcription, which aids hearers in disambiguating between discourse and sentential uses of cue phrases. Empirical evidence from more general studies of the intonational characteristics of word classes tends to support this possibility. Studies of portions of the London-Lund corpus such as Altenberg (1987) have provided intonational profiles of word classes including **discourse items**, conjunctions and adverbials which are roughly compatible with the notion that cue phrases tend to be deaccented, although the notion of discourse item used in this study is quite restrictive.[1] However, while the instance of *now* in Example (6) is in fact **reduced**, as Halliday and Hassan (1976) propose, that in Example (10), while interpreted as a discourse use, is nonetheless clearly intonationally prominent. Furthermore, both of the *now*s in Example (5) are also prominent. So it would seem that intonational prominence alone is insufficient to disambiguate between sentential and discourse uses.

In this paper we present a more complex model of intonational features and text-based features that can serve to disambiguate between sentential and discourse instances of cue phrases. Our model is based on several empirical studies (Hirschberg and Litman, 1987; Litman and Hirschberg, 1990): two studies of individual cue phrases in which we develop our model, and a more comprehensive study of cue phrases as a class, in which we confirm and expand our model. Before describing these studies and their results, we must first describe the intonational features examined in our analyses.

[1] In the 48 minute text Altenberg examines, he finds only 23 discourse markers, or about 17% of what our study of a similar corpus described in Section 5 would have predicted.

Figure 1: Deaccenting *now*.

3 Phrasing and Accent in English

The importance of intonational information to the communication of discourse structure has been recognized in a variety of studies (Butterworth, 1975; Schegloff, 1979; Brazil, Coulthard, and Johns, 1980; Hirschberg and Pierrehumbert, 1986; Pierrehumbert and Hirschberg, 1990; Silverman, 1987). However, just which intonational features are important and how they communicate discourse information is not well understood. Prerequisite, however, to addressing these issues is the adoption of a framework of intonational description to identify which intonational features will be examined and how they will be characterized. For the studies discussed below, we have adopted Pierrehumbert's (1980) theory of English intonation, which we will describe briefly below.

In Pierrehumbert's phonological description of English, intonational contours, or **tunes**, are described as sequences of **low** (**L**) and **high** (**H**) tones in the **fundamental frequency (F0) contour**, the physical correlate of pitch. These tunes have as their domain the **intonational phrase**, and are defined in terms of the **pitch accent**(s), **phrase accent**(s), and **boundary tone** which together comprise an intonational phrase.

One of the intonational features we examine with respect to cue phrases is the accent status of each cue, that is, whether or not the cue phrase is accented, or made intonationally prominent, and, if it is accented, what type of pitch accent it bears. Pitch accents usually appear as peaks or valleys in the F0 contour. They are aligned with the stressed syllables of lexical items, making those items prominent. Note that, while every lexical item in English has a lexically stressable syllable, which is the rhythmically most prominent syllable in the word, not every stressable syllable is in fact accented; so, lexical stress is distinguished from pitch accent. Lexical items which do bear pitch accents are said to be **accented**, while those not so marked are said to be **deaccented**. Items that are deaccented tend to be function words or items that are **given** in a discourse (Prince, 1981). For example, in Figure 1, *now* is deaccented, while *cue* is accented. Contrast Figure 1 with Figure 2. For ease of comparison, we present F0 contours of synthetic speech, where the x-axis represents time and the y-axis, frequency in Hz.[2] In Figure 1, the first F0 peak occurs on *let's*; in Figure

[2] The synthetic contours were synthesized by the Bell Labs Text-to-Speech System (Olive and Liberman, 1985) and displayed using WAVES speech analysis software (Talkin, 1989).

Figure 2: **H*** Accent on *now*.

2, the first peak occurred on *now*. The most prominent accent in a phrase is termed the **nuclear stress**, or **nuclear accent**, of the phrase. In both Figures 1 and 2, *cue* bears nuclear stress. In addition to the F0 excursions illustrated in Figures 1-5, accented syllables tend to be longer and louder than deaccented syllables, so there are a number of acoustic correlates of this perceptual phenomenon.

In Pierrehumbert's description of English, there are six types of pitch accent, all composed of either a single low (**L***) or high (**H***) tone or an ordered pair of low and high tones, such as **L+H*** or **H*+L**. In each case, the tone aligned with the stressed syllable of the accented lexical item is indicated by a star (*); thus, if *telephone* is uttered with a **L*+H** accent, the low tone (**L***) is aligned with the stressed syllable */tel/*, and the **H** tone falls on the remainder of the word. For simple pitch accents, of course, the single tone is aligned with the stress. The pitch accents in Pierrehumbert's description of English include two simple tones — **H*** and **L*** — and four complex ones — **L*+H**, **L+H***, **H*+L**, and **H+L***. The most common accent, **H***, comes out as a peak on the accented syllable (as, on *now* in Figure 2). **L*** accents occur much lower in the speaker's pitch range than **H*** and are phonetically realized as local F0 minima. The accent on *now* in Figure 3 is a **L***. Figure 4 shows a version of the sentence in Figures 1-3 with a **L+H*** accent on the first instance of *now*. Note that there is a peak on *now* (**H***) — as there was in Figure 2 — but now a striking valley (**L**) occurs just before this peak.

In Pierrehumbert and Hirschberg (1990), a compositional approach to intonational meaning is proposed in which pitch accents are viewed as conveying **information status**, such as newness or salience, about the denotation of the accented items and the relationship of denoted entities, states, or attributes to speaker and hearer's **mutual beliefs** about the discourse. In particular, it is claimed that speakers use **H*** accents to indicate that an item represents **new** information, which should be added to their mutual belief space. For example, standard declarative utterances in English commonly involve **H*** accents. **L*** accents, on the other hand, are used to indicate that an item is salient in the discourse but for some reason should not be part of what is added to the mutual belief space; standard yes-no question contour in English employs **L*** accents. The meanings associated with the **H+L** accents are explained in terms of the accented item's inferability from the mutual belief space: **H*+L** items are marked as inferable from the mutual belief space but nonetheless part of what is to be added to that space; **H+L*** accents are inferable and not to be added to speaker and hearer's mutual beliefs. **L+H** accents are defined in terms of the evocation of a **scale**, defined

Figure 3: **L*** Accent on *now*.

Figure 4: An **L+H*** Accent.

Figure 5: Two Phrases.

as a partially ordered set following (Hirschberg, 1991): **L*+H** accents, often associated with the conveyance of uncertainty or of incredulity, evoke a scale but predicate nothing of the accented item with respect to the mutual belief space; **L+H*** accents, commonly associated with contrastive stress, also evoke a scale but do add information about the accented item to speaker and hearer's mutual belief space (Pierrehumbert and Steele, 1987; Hirschberg and Ward, 1992).

Another intonational feature which is considered in our study of cue phrases is prosodic phrasing. There are two levels of such phrasing in Pierrehumbert's theory, the intonational phrase and the **intermediate phrase**, a smaller sub-unit. A well-formed intermediate phrase consists of one or more pitch accents plus a high (**H**) or low (**L**) phrase accent. The phrase accent controls the pitch between the last pitch accent of the current intermediate phrase and the beginning of the next — or the end of the utterance. An intonational phrase is composed of one of more intermediate phrases, plus a boundary tone. Boundary tones may be high (**H%**) or low (**L%**) also, and fall exactly at edge of the intonational phrase. So, each intonational phrase ends with a phrase accent and a boundary tone.

A given sentence may be uttered with considerable variation in phrasing. For example, the utterance in Figure 2 was produced as a single intonational phrase, whereas in Figure 5 *now* is set off as a separate phrase.

Intuitively, prosodic phrases divides an utterance into meaningful "chunks" of information (Bolinger, 1989). Variation in phrasing can change the meaning hearers assign to tokens of a given sentence. For example, the interpretation of a sentence like *Bill doesn't drink because he's unhappy* is likely to change, depending upon whether it is uttered as one phrase or two. Uttered as a single phrase, this sentence is commonly interpreted as conveying that Bill does indeed drink — but the cause of his drinking is not his unhappiness. Uttered as two phrases (*Bill doesn't drink — because he's unhappy*), it is more likely to convey that Bill does not drink — and the reason for his abstinence is his unhappiness. In effect, variation in phrasing appears to change the scope of negation in the sentence. When the sentence is uttered as a single phrase the negative is interpreted as having wide scope — over the entire phrase, and, thus, the entire sentence. When *Bill doesn't drink* is separated from the second clause by a phrase boundary, the scope of negation is limited to just the first clause.

The occurrence of phrase accents and boundary tones in the F0 contour, together with other phrase-final characteristics such as pause, decrease in amplitude, glottalization of phrase-final syllables, and phrase-final syllable lengthening, enable us to identify intermediate and intonational phrases in natural speech. Identification of pitch accents and phrase boundaries using a prosodic transcription system based on the one employed here has been found to be quite reliable between transcribers.[3]

Meaningful intonational variation has been found in studies of phrasing, choice of accent type and location, overall tune type, and variation in **pitch range**, where the pitch range of an intonational phrase is defined by its **topline**, roughly, the highest peak in the F0 contour of the phrase, and the speaker's **baseline**, the lowest point the speaker realizes in normal speech, measured across all utterances. In the studies described below, we examined each of these features, in addition to text-based features, to see which best predicted cue phrase disambiguation, and to look for associations among text-based and intonational features.

4 Single Cue Phrase Studies

Our first study of cue phrase disambiguation investigated multi-speaker usage of the cue phrase *now* in a recorded, transcribed radio call-in program (Hirschberg and Litman, 1987). Our corpus consisted of four days of the radio call-in program "The Harry Gross Show: Speaking of Your Money", recorded during the week of 1 February 1982 (Pollack, Hirschberg, and Webber, 1982). In this Philadelphia program, Gross offered financial advice to callers; for the 3 February show, he was joined by an accountant friend, Fred Levy. The four shows provided approximately ten hours of conversation between expert(s) and callers. The corpus was transcribed by Martha Pollack and Julia Hirschberg in 1982, in connection with another study.

We chose *now* for this initial study for several reasons. First, the corpus contained numerous instances of both discourse and sentential usages of *now* (approximately 350 in all). Second, *now* often appears in conjunction with other cue phrases, e.g. *well now*, *ok now*, *right now*. This allowed us to study how adjacent cue phrases interact with one another. Third, *now* has a number of desirable phonetic characteristics. As it is monosyllabic, possible variation in stress patterns do not arise to complicate the analysis. Because it is completely voiced and introduces no segmental effects into the F0 contour, it is also easier to analyze pitch tracks reliably.

Our model was initially developed from a sample consisting of 48 occurrences of *now* — all the instances from two sides of tapes of the show chosen at random. Two instances were excluded since the phrasing was difficult to determine due to hesitation or interruption. To test the validity of our initial hypotheses, we then replicated our study with a second sample from the same corpus, the first 52 instances of *now* taken from another four randomly chosen sides of tapes. We excluded two tokens from these tapes because of lack of available information about phrasing or accent and five others because we were unable to decide whether the tokens were discourse or sentential.

Our data analysis included the following steps. First, the authors determined separately, and by ear, whether individual tokens were discourse or sentential usages and tagged the transcript of the corpus accordingly. We then digitized and pitch-tracked the intonational phrase containing each token, plus the preceding and succeeding intonational phrases, if produced by the same speaker.[4] Intonational features were determined by one of the authors from the speech and pitch tracks, separately from the discourse/sentential

[3] See results of several prosodic labeling experiments using TOBI, the TOnes and Break Indices transcription system (Silverman et al., 1992b; Silverman et al., 1992a).

[4] The pitch tracks in the first two studies were produced with a pitch tracker written by Mark Liberman. For the third study we used a pitch tracker written by David Talkin and WAVES speech analysis software (Talkin, 1989) in our prosodic analysis.

judgment. Discourse and sentential uses were then compared along several dimensions:

1. Each instance of *now* was examined to determine if it was accented and, if so, to determine what type of accent was employed.
2. Differences in phrasing, in particular whether or not *now* represented an entire intermediate or intonational phrase, were identified.
3. *Now*'s position in its intonational and its intermediate phrase (first, not first but preceded only by other cue phrases, last, or none of these) was noted.
4. The type of intonational contour used over the phrase in which *now* occurred was determined.
5. Whether and how *now* occurred adjacent to other cue phrases was noted.
6. The position of the phrase containing *now* with respect to speaker turn was noted.

Of these comparisons, the first three turned out to distinguish between discourse and sentential *now* quite reliably. In particular, a combination of accent type, phrasal composition, and phrasal position reliably distinguished between the tokens in the corpus.

4.1 Results of Intonational Analysis

Of the 100 tokens of *now* from the combined 48- and 52-token corpora, just over one-third of our samples (37) were judged to be sentential, and just under two-thirds (63) discourse. The first striking difference between the two appeared in the composition of the intermediate phrase containing the item, as illustrated in Table 1. Of all the sentential uses of *now*, only one appeared as the only item in an intermediate phrase,

Table 1: Phrasing for *now*, N=100.

	Part of Larger Intermediate Phrase	Alone in Intermediate Phrase
Sentential	36	1
Discourse	37	26

while 26 (41.3%) discourse *now*s represented entire intermediate phrases. Of these 26, one half constituted the only lexical item in a full intonational phrase. So, our findings suggested that *now* set apart as a separate intermediate phrase is very likely to be interpreted as conveying a discourse meaning rather than a sentential one.

Another clear distinction between discourse and sentential *now* emerged when we examined the surface position of *now* within its intermediate phrase. As Table 2 illustrates, 62 of the 63 discourse *now*s (98.4%) were **first-in-phrase**, absolutely first or followed only another cue phrase in their intermediate phrase; of these, 59 (95.2%) were also absolutely first in their intonational phrase, that is, first in major prosodic phrase and *not* preceded by any other cue phrases. Only five (13.5%) sentential tokens were first-in-phrase. Also, while 22 (59.5%) sentential *now*s were phrase final, only one discourse token was so positioned. So, once

Table 2: Position Within Intermediate Phrase for *now*, N=100.

	First	Last	Other
Sentential	5	22	10
Discourse	62	1	0

intermediate phrases are identified, discourse and sentential *now* appear to be generally distinguishable by position within the phrase.

Finally, discourse and sentential occurrences were distinguishable in terms of presence or absence of pitch accent — and by type of pitch accent, where accented. Because of the large number of possible accent types, and since there are competing reasons to accent or deaccent items, such as accenting to indicate **contrastive stress** or deaccenting to indicate an item is already given in the discourse, we might expect these findings to be less clear than those for phrasing. In fact, although their interpretation is more complicated, the results are equally striking.

Results of an analysis of the 97 occurrences from this sample for which accent type could be precisely determined are presented in Table 3. Of those tokens not included, two discourse tokens were judged either **L*** or **H*** with a compressed pitch range, and one discourse token was judged either deaccented or **L***. Note

Table 3: Accenting of Discourse and Sentential *now*, N=97.

	Deaccented	H*orComplex	L*
Sentential	5	32	0
Discourse	31	16	13

first that large numbers of discourse and sentential tokens were uttered with a **H*** or complex accent — 16 (26.7%) discourse and 32 (86.5%) sentential tokens. The chief similarity here lies in the use of the **H*** accent type, with 14 discourse uses and 14 sentential; seven other sentential tokens are ambiguous between **H*** and complex. Note also that discourse *now* was much more likely overall to be deaccented — 31 of the 60 discourse tokens (51.7%) versus 5 of the 37 sentential *now*s (13.5%). No sentential *now* was uttered with a **L*** accent — although 13 discourse *now*s were.

An even sharper distinction in accent type is found if we exclude those *now*s which are alone in intermediate phrase from the analysis. Recall from Table 1 that all but one of these tokens represented a discourse use. These *now*s were always accented, since it is generally the case that each intermediate phrase contains at least one pitch accent. Of the discourse tokens representing entire intermediate phrases for which we can distinguish accent type precisely, 14 bore **H*** accents. This suggests that one similarity between discourse and sentential *now* — the frequent **H*** accent — might disappear if we limit our comparison to those tokens forming part of larger intonational phrases. In fact, such is the case, as is shown in Table 4.

The majority, 31 (86.1%), of sentential *now*s forming part of larger intonational phrases received a **H*** or complex pitch accent, while all 36 discourse *now*s forming part of larger intonational phrases were deaccented or bore a **L*** accent. In fact, those discourse *now*s not distinguishable from sentential by being set apart as separate intonational phrases were generally so distinguishable with respect to pitch accent. Of the

Table 4: Accenting of *now* in Larger Intonational Phrases, N=72.

	Deaccented	H*orComplex	L*
Sentential	5	31	0
Discourse	31	0	5

three discourse tokens whose pitch accent type was not identifiable, which were omitted from Table 3, two were set apart as separate intonational phrases and one was judged either to bear a **L*** pitch accent or to be deaccented. Thus, all three could be distinguished from sentential tokens in terms of accent type and phrasing. Furthermore, of the five deaccented sentential *now*s in Table 4, none were first-in-phrase, while only one of the deaccented discourse tokens was similarly non-initial. In fact, of the 100 tokens in our initial study of *now*, all but two were distinguishable as discourse or sentential in terms of a combination of position in phrase, phrasal composition, and accent.

Thus, we were able to hypothesize from our study of *now* that discourse uses were either uttered as a single intermediate phrase or in a phrase containing only cue phrases (Discourse Type A), or uttered at the beginning of a longer intermediate phrase or preceded only by other cue phrases in the phrase and with a **L*** pitch accent or without a pitch accent (Discourse Type B).[5] Only one of the 37 cue phrases judged to be of Sentential Type was uttered as a single phrase. If first-in-phrase, they were nearly always uttered with a **H*** or complex pitch accent (Sentential Type A); if not first-in-phrase, they could bear any type of pitch accent or be deaccented (Sentential Type B). These results are summarized in Figure 6.

Figure 6: Prosodic Characteristics of Discourse and Sentential Uses.

Cue Phrases

Sentential

A: Initial in Larger Phrase, **H** or Complex Accent

B: Non-Initial in Larger Phrase, Accented or Deaccented

Discourse

A: Alone in Phrase, Accented or Deaccented

B: Initial in Larger Phrase, Deaccented or **L** Accent

[5] We also investigated whether the different prosodic models of discourse uses could be mapped to the different meanings that discourse uses can convey — as discussed in Section 1 and illustrated in Table 14 — but found no evidence for such a mapping in our data. However, other authors have found more promising results for the cue phrase *ok* (Swora and Beckman, 1991; Hockey, 1991).

4.2 Speaker Variability

Since the preponderance of tokens in our sample from one professional speaker might well skew our results, we compared characteristics of phrasing and accent for host and non-host data. The results showed no significant differences between host and caller tokens in terms of the hypotheses proposed above. First, host (n=37) and callers (n=63) produced discourse and sentential tokens in roughly similar proportions — 40.5% sentential for the host and 34.9% for his callers. Similarly, there was no distinction between host and non-host data in terms of choice of accent type, or accenting versus deaccenting. Our findings for position within phrase also hold for both host and non-host data. However, in tendency to set discourse *now* apart as a separate intonational or intermediate phrase, there was an interesting distinction. While callers tended to choose from among the two options for discourse *now* in almost equal numbers (48.8% of their discourse *now*s were separate phrases), the host chose this option only 27.3% of the time. However, although host and caller data differed in the proportion of occurrences of the two classes of discourse *now* which emerge from our data as a whole, the existence of the classes themselves was confirmed. Where the host did not produce discourse *now*s set apart as separate intonational or intermediate phrases, he always produced discourse *now*s which were deaccented or accented with a **L*** accent. We hypothesize, then, that, while individual speakers may choose different strategies to realize discourse *now*, they appear to choose from among the same limited number of options.

4.3 Distinguishing Discourse and Sentential Usage in Transcriptions

Our conclusion from this study, that intonational features play a crucial role in the distinction between discourse and sentential usage in speech, clearly poses problems for text. Do readers use strategies different from hearers to make this distinction, and, if so, what might they be? Are there perhaps orthographic correlates of the intonational features which we have found to be important in speech? As a first step toward resolving these questions, we examined the orthographic features of the transcripts of our corpus, which, as noted on page 10, had been prepared independently of this study and without regard for intonational analysis.

We examined transcriptions of all tokens of *now* in our combined sample to determine whether prosodic phrasing was reliably associated with orthographic marking. There were no likely orthographic clues to accent type or placement, such as capitalization, in the transcripts. Of all 60 instances of *now* that were absolutely first in their intonational phrase, 34 (56.7%) were preceded by punctuation — a comma, dash, or end punctuation — and 17 (28.3%) were first in speaker turn, and thus orthographically marked by indication of speaker name. So, in 51 (85%) cases, first position in intonational phrase coincided with orthographic indicators in the transcript. No *now*s that were *not* absolutely first in their intonational phrase — for example, none that were merely first in their intermediate phrase — were so marked. Of those 23 *now*s coming last in an intermediate or intonational phrase, however, only 14 (60.9%) were immediately followed by a similar orthographic clue. Finally, of the 13 instances of *now* which formed separate intonational phrases, only two (15.4%) were distinguished orthographically by being both preceded and followed by some orthographic indicator. And none of the *now*s that formed complete intermediate phrases, but *not* complete intonational phrases, were so marked.

These findings suggest that, of the intonational features we found useful in disambiguating cue phrases in speech, only the feature first in intonational phrase has any clear orthographic correlate. This correlation, however, seems potentially to be a useful one. Of the 63 discourse *now*s in our corpus, recall that 59 (93.7%)

were first in their intonational phrase. Of these 59, 48 were preceded by orthographic indicators in the transcription, as described above. Of sentential cues, 22 were last in their intermediate phrase, and, of these, 13 were followed by some orthographic indicator in the transcription. Of 34 cue phrases that were neither preceded nor followed by orthographic markings in the transcription, the majority (21, or 61.8%) were sentential uses. If we predict sentential/discourse usage based simply on the presence or absence of preceding and succeeding orthographic markings, we would predict that cue phrases preceded by orthographic indicators represent discourse uses, and that phrases either followed by orthographic indicators or neither preceded nor followed would be sentential uses, for a total of 82 correct predictions for the 100 cue phrases in this study. Thus, 82% of *now*s might be orthographically distinguished. We will have more to say on the role of orthography in disambiguating cue phrases in connection with the study described in Section 5.

4.4 Multi-Speaker Study of *Well*

Based on the findings of our study of *now*, we proposed that listeners may use prosodic information to disambiguate discourse from sentential uses of cue phrases (Hirschberg and Litman, 1987). However, although we chose to study *now* for its ambiguity between discourse and sentential (temporal adverbial) uses, it may of course also be seen as representative of sense ambiguities between temporals and non-temporals or deictics and non-deictics. Thus, if indeed our findings generalize, it might be to a class we had not intended to investigate. To discover further evidence that our results did indeed apply to the discourse/sentential use disambiguation, we conducted another multi-speaker study, this time of the discourse and sentential uses of the single cue phrase *well*. Again, our corpus consisted of recordings of the Harry Gross radio call-in program. In addition, we used tokens from several other corpora of recorded, transcribed speech, including the corpus described in Section 5. This time we included no more than three tokens from any speaker to minimize the potential effect of speaker idiosyncracy.

Our findings for this study of *well* were almost identical to results from the earlier study of *now*, described above. Briefly, of the 52 instances of *well* we examined, all but one token fit the model constructed from the results of the *now* study, depicted in Figure 6. In particular, of the 25 sentential uses of *well*, none constituted a single intermediate or intonational phrase. Only two sentential tokens were first-in-phrase, and both of these bore **H*** pitch accents. However, of the 27 discourse tokens of *well*, 14 were indeed alone in their intonational or intermediate phrases. All of the remaining 13 occurred first-in-phrase, and, of these 12 were deaccented. In all, 51 (98.1%) of the tokens in this study fit our model; the single counter-example was one discourse token, which bore a **H*** pitch accent and was part of a larger phrase.

Our study of *well* thus appeared to confirm our earlier results, and, in particular, to lend support to our hypothesis that cue phrases can be distinguished intonationally. However, although we had shown that two cue phrases appeared to pattern similarly in this respect, we had still not demonstrated that our model could be extended to cue phrases in general. To address this larger issue, we next conducted a single-speaker multi-cue phrase study.

5 The Single-Speaker/Multi-Cue Phrase Study

In this study, we examined all cue phrases consisting of a single lexical item that were produced by one speaker during 75 minutes, approximately 12,500 words, of recorded speech. Results of a pilot study of this corpus are reported in (Litman and Hirschberg, 1990). We limited ourselves here to the examination of single lexical items, since the hypothesis we had previously developed applies only to such items; e.g. it would be meaningless to ask whether a larger phrase bears a pitch accent or not. The corpus consisted of a keynote address given from notes by Ronald Brachman at the *First International Conference on Expert Database Systems* in 1986. This talk yielded 953 tokens, based upon a set of possible cue phrases derived from (Cohen, 1984; Grosz and Sidner, 1986; Litman and Hirschberg, 1990; Reichman, 1985; Schiffrin, 1987; Warner, 1985; Zuckerman and Pearl, 1986). The frequency distribution of the tokens is shown in Table 5.

Table 5: Distribution of Cue Phrases (N=953).

Cue Phrase	Tokens	Cue Phrase	Tokens
actually	32	*next*	4
also	9	*no*	9
although	8	*now*	69
and	320	*ok*	6
basically	5	*or*	63
because	12	*otherwise*	2
but	61	*right*	7
essentially	2	*say*	35
except	3	*second*	3
finally	11	*see*	26
first	21	*similarly*	5
further	11	*since*	2
generally	7	*so*	60
however	8	*then*	13
indeed	9	*therefore*	2
like	61	*well*	29
look	35	*yes*	3

By far the most frequent cue phrase occurring in our corpus is the conjunction *and*, representing 320 (33.6%) tokens. The next most frequent item is *now*, with only 69 occurrences. Other items occurring more than fifty times each in the corpus are *but*, *like*, *or*, and *so*. Note that there are 444 conjunctions — *and*, *but*, and *or* — comprising nearly half of the cue phrases in our corpus. In addition to the items shown in Table 5, we searched the corpus unsuccessfully for instances of the following cue phrases proposed in the literature (cf. Table 14): *accordingly*, *alright*, *alternately*, *alternatively*, *altogether*, *anyway*, *boy*, *consequently*, *conversely*, *fine*, *furthermore*, *gee*, *hence*, *hey*, *incidentally*, *likewise*, *listen*, *meanwhile*, *moreover*, *namely*, *nevertheless*, *nonetheless*, *nor*, *oh*, *though*, *yet*.

However, note that the set of items included in Table 14 is not identical to the set we have considered in this paper. In particular, we do consider the items *actually*, *basically*, *essentially*, *except*, *generally*, *no*, *right*, *since*, and *yes* (cf. Table 5), although they are not considered in the studies included in Table 14. We do not consider *again*, *equally*, *hopefully*, *last*, *only*, *overall*, *still*, *thus*, *too*, *unless*, *where*, *whereas*, and *why*, although these have been included by others in the set of possible cue phrases.

The temporal pattern of cue phrase use in the corpus exhibits some interesting features. While tokens were distributed fairly evenly during the middle portion of the talk, the first and last portions were less regular. The first decile of the transcript, defined by length in words, contained 140 cue phrases (14.7%), a higher proportion than any other decile of the corpus, while the second decile contained only 73 (7.7%). And the last decile of the talk contained an even lower proportion of cue phrases — only 64 (6.7%). So, it appears that, at least for this genre, cue phrases occur more frequently in the introductory remarks, and less frequently in the conclusion.

To classify each token as discourse or sentential, the authors separately judged each one by ear from the taped address while marking a transcription. Where we could not make a decision, we labeled the token ambiguous; so, any token could be judged "discourse", "sentential", or "ambiguous". The address was transcribed independently of our study by a member of the text processing pool at AT&T Bell Laboratories. In examining the transcription, we found that 39 cue phrases had been omitted by the transcriber: one token each of *actually*, *essentially*, *or*, and *well*, three tokens each of *so* and *ok*, nine tokens of *and*, and twenty tokens of *now*. It seemed significant that all but five of these were subsequently termed discourse uses by both judges — that is, that discourse uses seemed somehow omissible to the transcriber. One of the authors then assessed each tokens prosodic characteristics, as described on page 11.

In examining our classification judgments, we were interested in areas of disagreement as well as agreement. The set of tokens whose classification we both agreed upon and found unambiguous provided a testbed for our investigation of the intonational features marking discourse and sentential interpretation. The set of tokens one or both of us found ambiguous we examined to determine how intonation might in fact have contributed to that ambiguity. Table 6 presents the distribution of our judgments, where **classifiable** includes those tokens we both assigned either discourse or sentential, **ambiguous** identifies those we both were unable to classify, **partial disagreement** includes those only one of us was able to classify, and **complete disagreement** represents those tokens one of us classified as discourse and the other as sentential. Of the

Table 6: Judgments for All Tokens and for Conjunctions Alone (N=953).

Type	Total	Agreements		Disagreements	
		Classifiable	Ambiguous	Partial	Complete
All	953	878	59	11	5
Conjuncts	444	383	48	9	4
Non-Conjuncts	509	495	11	2	1

953 tokens in this corpus, we agreed in our judgments of 878 cue phrases (92.1%) as discourse or sentential. Another 59 (6.2%) tokens we both judged ambiguous. We disagreed on only 16 items (1.7%); for 11 of these, the disagreement was between classifiable and ambiguous.

When we examined the areas of ambiguity and disagreement in our judgments, we found that a high proportion of these involved judgments of coordinate conjunction tokens, *and*, *or*, and *but*, which, as we previously noted, represent nearly half of the tokens in this study. Table 6 shows that, comparing conjunction with non-conjunction, we agreed on the classification of 495 (97.2%) non-conjunction tokens but only 383 (86.3%) conjunctions. We both found 48 (10.8%) conjunctions ambiguous, but only 11 (2.2%) non-conjunctions; 48 of the 59 tokens we agreed were ambiguous in the corpus were, in fact, coordinate conjunctions. Of the 16 tokens on which we simply disagree, 13 (81.3%) were conjunctions.

The fact that conjunctions account for a large number of the ambiguities we found in the corpus and the disagreements we had about classification is not surprising when we note that the discourse meanings of conjunction as described in the literature (see Table 14) seem to be quite similar to the meanings of sentential conjunction. For example, the discourse use of *and* is defined as 'parallelism' in (Cohen, 1984), 'a marker of addition' or 'sequential continuity' in (Schiffrin, 1987), and 'conjunction' in (Warner, 1985). These definitions fail to provide clear guidelines for distinguishing discourse uses from sentential, as in cases like Example (11) (RJB86). Here, while the first *and* seems intuitively sentential, the second is much more problematic.

(11) But instead actually we are bringing some thoughts on expert databases from a place that is even stranger *and* further away *and* that of course is the magical world of artificial intelligence.

However, while similarities between discourse and sentential interpretations appear to make conjunction more difficult to classify than other cue phrases, the same similarities may make the need to classify them less important from either a text generation or a text understanding point of view.

Once we had classified the tokens in the corpus, we analyzed them for their prosodic and syntactic features as well as their orthographic context, in the same way we had examined tokens for the earlier two studies. In each case, we noted whether the cue phrase was accented or not and, if accented, we noted the type of accent employed. We also looked at whether the token constituted an entire intermediate or intonational phrase — possibly with other cue phrases — or not, and what each token's position within its intermediate phrase and larger intonational phrase was — first-in-phrase (again, including tokens preceded only by other cue phrases as well as tokens that were absolutely first in intermediate phrase), last, or other. We also examined each item's part-of-speech, using Church's (1988) part-of-speech tagger. Finally, we investigated orthographic features of the transcript which might be associated with a discourse/sentential distinction, such as immediately preceding and succeeding punctuation and paragraph boundaries. In both the syntactic and orthographic analyses we were particularly interested in discovering how successful non-prosodic features which might be obtained automatically from a text would be in differentiating discourse from sentential uses.

5.1 Results of the Intonational Analysis

We looked first at the set of 878 tokens whose classification as discourse or sentential we both agreed upon. Our findings from this set confirmed the prosodic model we found in the studies described above to distinguish discourse from sentential uses successfully. The distribution of these judgments with respect to the prosodic model of discourse and sentential cue phrases depicted in Figure 6 is shown in Table 7. Recall that the

Table 7: Prosody of Classified Tokens (N=878).

Judgment	Prosody	
	Discourse	Sentential
Discourse	301	40
Sentential	176	361

(χ^2 = 258.863, df = 1, p <.001)

prosodic model in Figure 6 includes the following intonational profiles: Discourse Type A, in which a cue phrase constitutes an entire intermediate phrase, or is in a phrase containing only other cue phrases, and may have any type of pitch accent; Discourse Type B, in which a cue phrase occurs at the beginning of a larger intermediate phrase, or is preceded only by other cue phrases, and bears a **L*** pitch accent or is deaccented; Sentential Type A, in which the cue phrase occurs at the beginning of a larger phrase and bears a **H*** or complex pitch accent; and Sentential Type B, in which the cue phrase occurs in non-initial position in a larger phrase. Table 7 shows that our prosodic model fits the new data reasonably well, successfully predicting 662 (75.4%) of the classified tokens. Of the 341 cue phrases we both judged discourse, 301 (88.3%) fit the prosodic discourse model; 50 of these were of Discourse Type A and 251 were of Discourse Type B. Of the 537 tokens we both judged sentential, 361 (67.2%) fit one of the prosodic sentential models. The overall ratio of cue phrases judged discourse to those judged sentential was about 2:3. A χ^2 test shows significance at the .001 level.[6] While these results are highly significant, they clearly do not match the previous findings for *now* and *well* discussed in Section 4, in which all but three tokens fit our model.

So, for this larger study, the tokens which did **not** fit our prosodic model remain to be explained. In fact, there is some regularity among these counter-examples. For example, eight (20%) of the items judged discourse that did not fit our discourse prosodic model were tokens of the cue phrase *say*. All of these failed to fit our prosodic discourse model by virtue of the fact that they occurred in non-initial phrasal position; such items are illustrated in Example (8). Of the 176 items judged sentential which failed to fit our sentential prosodic model, 138 (78.4%) were conjunctions. Eleven of these fit the Discourse Type A prosodic model and 127 fit the Discourse Type B model. Both judges found such items relatively difficult to distinguish between discourse and sentential use, as discussed above. Table 8 shows how judgments are distributed with respect to our prosodic model when coordinate conjunctions are removed from the sample. Our model thus predicts

Table 8: Prosody of Classified Non-Conjuncts (N=495).

Judgment	Prosody	
	Discourse	Sentential
Discourse	167	35
Sentential	38	255

(χ^2 = 239.43, df = 1, p <.001)

422 (85.3%) of non-conjunction cue phrase distinctions, somewhat better than the 662 (75.4%) successful predictions for all classified cue phrases, as shown in Table 7.

Our prosodic model itself can of course be decomposed to examine the contributions of individual features to discourse/sentential judgments. Table 9 shows the distribution of judgments by all possible feature complexes for all tokens. Note that four cells (ONFD, ONFH, ONFL, and ONFC) are empty, since all items alone in their intermediate phrase must perforce come first in it.

This distribution reveals that there is considerable agreement when cue phrases appear alone in their intermediate phrase (tokens coded with initial OF, corresponding to Discourse Type A in Figure 6): such

[6] The χ^2 test measures the degree of association between two variables by calculating the probability (p) that the disparity between expected and actual values in each cell is due to chance. The value of χ^2 itself for n degrees of freedom (df) is an overall measure of this disparity.

Table 9: Prosodic Feature Configurations and Judgments (N=953).

Model	Code	Tokens	Judgments		Unclassifiable	
			% Discourse	% Sentential	%	Tokens
Discourse A	OFD	7	42.86	42.86	14.29	1
Discourse A	OFH	35	68.57	25.71	5.71	2
Discourse A	OFL	106	82.08	8.49	9.43	10
Discourse A	OFC	28	92.86	7.14	0	0
	ONFD	NA	NA	NA	NA	NA
	ONFH	NA	NA	NA	NA	NA
	ONFL	NA	NA	NA	NA	NA
	ONFC	NA	NA	NA	NA	NA
Discourse B	NOFD	307	42.35	44.30	13.36	41
Discourse B	NOFL	55	56.36	30.91	12.73	7
Sentential A	NOFH	42	19.05	69.05	11.90	5
Sentential A	NOFC	40	42.50	52.50	5.00	2
Sentential B	NONFD	154	1.30	95.45	3.25	5
Sentential B	NONFL	18	50.00	44.44	5.60	1
Sentential B	NONFC	58	0	100.00	0	0
Sentential B	NONFH	103	3.88	95.15	.97	1

Feature complexes are coded as follows:
Initial O or NO – consists of a single intermediate phrase or not.
Medial F or NF – appears first-in-phrase or not.
Final D,H, L, or C – deaccented, or bears a **H***, **L*** or complex pitch accent.

items are most frequently judged to be discourse uses. There is also considerable agreement (163 tokens, or 92.6%) on the classification of the tokens between the authors in such cases.

There is even greater agreement when cue phrases appear in non-initial position in a larger intermediate phrase (NONF* — Sentential type B in Figure 6); these tend to be judged sentential. When the token is deaccented, or receives a complex or high accent (NONFD, NONFC and NONFH), the fit with the model, as well as the agreement figures on classification, are especially striking. A small majority of tokens in the **L*** accent class (NONFL) do not fit the sentential prosodic model; note that the agreement level producing this classification was good. However, as with the OFD subtype of Discourse Type A, which also has the worst results for its class, we have the fewest tokens for this prosodic type.

Tokens which fit Discourse type B in Figure 6 — first in a larger phrase and deaccented (NOFD) or first in a larger phrase and bearing a **L*** accent (NOFL) — appear more problematic: of the former, there was more disagreement than agreement between the judge's classification and the prosodic prediction of the classification. And of the 153 sentential items that fit this discourse prosodic model, 127 (83.0%) are conjunctions. The level of disagreement for the judge's classifications was also highest for Discourse type B.

While there is more agreement that tokens corresponding to Sentential Model A and characterized as NOFH — first in a larger phrase with a **H*** accent — or NOFC — first in a larger phrase and bearing a complex pitch accent, are sentential, this agreement is certainly less striking than in the case of tokens corresponding to Sentential Model B and characterized here as NONF* — non-initial in a larger phrase with any type of pitch accent. Since Discourse Type B and Sentential Type A differ from each other only in type of pitch accent, we might conclude that the pitch accent feature is not as powerful a discriminator as the the fact that a potential cue phrase is alone in its intermediate phrase or first-in-phrase.

Finally, Table 10 presents a breakdown by lexical item of some of the data in Table 9. In this table we show the prosodic characteristics of classified cue phrases, indicating the number of items that fit our prosodic models and which models they fit, and the number that did not. First note that some cue phrases in our single-speaker study were always identified as sentential: *actually, also, because, except, first, generally, look, next, no, right, second, see, since, therefore,* and *yes*. A few were only identified as discourse: *finally, however,* and *ok*. In Section 4.2 we examined the possibility that different speakers might favor one prosodic strategy for realizing discourse or sentential usage over another, based on the data used in our study of *now*. Overall, the speaker in RJB86 favored the prosodic model Discourse B over Discourse A for cue uses in 251 (83.4%) cases. For sentential uses, this speaker favored the Sentential A model slightly over Sentential B, employing the former in 204 (56.8%) of cases. However, it is also possible that a speaker might favor prosodic strategies that are specific to particular cue phrases to convey that they are discourse or sentential. For example, from Table 10, we see that most discourse uses of all coordinate conjunctions fit our prosodic model Discourse B, while all occurrences of *finally* and *further* fit Discourse A. Of cue phrases classified as sentential, *actually, first, look, right, say, see, so, well* (and others) most frequently fit Sentential A, while *and* most frequently fits Sentential B.

5.2 Distinguishing Discourse and Sentential Usage in Transcriptions

As in our previous studies, we also examined potential non-prosodic distinctions between discourse and sentential uses. Of the orthographic and syntactic features we examined, we found presence or absence of preceding punctuation and part-of-speech to be most successful in distinguishing discourse from sentential uses. We also examined how and when cue phrases occurred adjacent to other cue phrases. Although the data are sparse — only 118 (12.4%) of our tokens occurred adjacent to other cue phrases, they suggest that

Table 10: Classified Cue Phrases by Prosodic Models (N=878).

Word	Fitting Prosodic Models				Not Fitting
	Discourse		Sentential		Models
	A	B	A	B	
actually			20	8	0
also			3	1	5
although		5	1		2
and	2	91	11	78	94
basically	1		3		1
because				3	5
but	2	23	1	2	24
essentially					0
except			1		2
finally	7				4
first			18	2	4
further	6		2	1	2
generally			5		1
however	3	2			3
indeed	2		2	1	3
like		2	20	27	9
look			30	3	2
next			2	2	0
no			5	2	2
now	8	50	6	3	1
ok	3	3			0
or	4	12	5	9	25
otherwise					1
right			6	1	0
say	1	16	9	1	8
second			3		0
see			22	4	0
similarly	2		1		2
since				1	1
so	2	39	9	4	6
then	2	1	1		9
therefore			2		0
well	5	7	15	2	0
yes			1		2
Total	50	251	204	155	218

co-occurrence data may provide information useful for cue phrase disambiguation. In particular, of the 26 discourse usages of cue phrases preceded by other classifiable cue phrases, 20 (76.9%) were also discourse usages. Similarly, out of 29 sentential usages preceded by a classified cue, 21 (72.4%) were preceded by another sentential use. With respect to classified cue phrases that were followed by other classified cue phrases, 20 out of 28 (71.4%) discourse usages were followed by a discourse usage, while 21 out of 27 (77.8%) sentential usages were followed by other sentential uses.

Table 11 presents the orthography found in the transcription of the cue phrases present in the recorded speech. The orthographic markers used by the transcriber include commas, periods, dashes, and paragraph

Table 11: Transcribed Classified Cue Phrases Associated with Orthography (N=843).

Position	Judgment	
	Discourse	Sentential
Preceding (only)	151	37
Succeeding (only)	12	21
Preceding and Succeeding	25	0
None	119	478

breaks. For the 843 tokens — 536 judged sentential and 307 judged discourse — whose classification both judges agreed upon, and excluding those items which the transcriber omitted, orthography or its absence is a useful predictor of discourse or sentential use. In particular, of the 213 tokens preceded by punctuation (combining rows one and three from Table 11), 176 (82.6%) are discourse usages. Note, however, that many discourse usages are *not* marked by preceding orthography; the 176 marked tokens represent only 57.3% of all discourses uses in this sample. Only 37 (6.9%) of sentential usages were also preceded by orthographic indicators. Twelve tokens that are succeeded but not preceded by orthographic markings are discourse and 21 are sentential. All of the tokens in RJB86 that are both preceded and succeeded by orthography are discourse usages, although, again, these 25 tokens represent only 8.1% of the discourse tokens in the sample. So, the presence of preceding orthographic indicators — especially in conjunction with succeeding indicators — appears to be a reliable textual indicator that a potential cue phrase should be interpreted as a discourse use, predicting correctly in 176 (82.6%) cases. While we found that discourse uses are not always reliably marked by such indicators in the RJB86 transcription, it is possible to predict the discourse/sentential distinction from orthography alone for this corpus in 675 (80.1%) cases.

In our study of *now*, described in Section 4.3, we found that, in 51 (85%) cases, cue phrases that were first in intonational phrase were marked orthographically. In the current single-speaker study, first position in intonational phrase was orthographically marked in only 199 of 429, or 46.4% of cases. So, in this study, the association between position in intonational phrase and orthographic marking appears much weaker.

We also found that part-of-speech could be useful in distinguishing discourse from sentential usage — although less useful than orthographic cues — as shown in Table 12.[7] If we simply predict discourse or sentential use by the assignment most frequently associated with a given part-of-speech, Church's part-of-speech algorithm predicts discourse or sentential use in 561 (63.9%) cases for tokens where both judges

[7] The part-of-speech tagger employed in this analysis (Church, 1988) uses a subset of the part-of-speech tags used in (Francis and Kucera, 1982). We have translated these for Table 12. Note that "intensifier" corresponds to "QL" in (Francis and Kucera, 1982)

Table 12: Part-of-Speech Analysis of Classified Cue Phrases (N=878).

Part-of-Speech	Judgement	
	Discourse	Sentential
article	0	6
coordinating conjunction	139	244
cardinal numeral	0	21
subordinating conjunction	43	58
preposition	0	3
adjective	1	12
singular or mass noun	10	7
singular proper noun	5	1
intensifier	4	6
adverb	118	101
verb, base form	21	78

agreed on discourse/sentential assignment. For example, we assume that since the majority of conjunctions and verbs are judged sentential, these parts-of-speech are predictors of sentential status, and, since most adverbials are associated with discourse uses, these are predictors of discourse status, and so on.

If we employ both orthographic indicators and part-of-speech as predictors of the discourse/sentential distinction, we achieve only slightly better prediction than with orthographic cues alone. That is, if we consider both an item's part-of-speech tag and adjacent orthographic indicators, we model the RJB86 data only marginally more accurately. Table 13 models correctly 677 (80.3%) transcribed, classified tokens in RJB86 from orthographic and part-of-speech information. For example, given a coordinating conjunction, our model would predict that it would be a discourse use if preceded by orthography, and a sentential use otherwise. In fact, the only difference from orthography alone is the way succeeding orthography can signal a discourse use for a singular or mass noun, and a sentential use for adverbs.

While the use of orthographic and part-of-speech data represents only a fractional improvement over orthographic information alone, it is possible that, since the latter is not subject to transcriber idiosyncracy, such an approach may prove more reliable than orthography alone in the general case. And, for text-to-speech applications, it is not clear how closely orthographic conventions for unrestricted written text will approximate the regularities we have observed in our transcribed corpora.

5.3 Summary

Our findings for our single-speaker multi-cue phrase study support the intonational model of discourse/sentential characteristics of cue phrases which we proposed based on our earlier multi-speaker single-cue phrase studies of *now* and *well* (Hirschberg and Litman, 1987; Litman and Hirschberg, 1990). In each study, discourse uses of cue phrases fit one of two prosodic models: in one, the cue phrase was set apart as a separate intermediate phrase, possibly with other cue phrases; in the other, the cue phrase was first-in-phrase, possibly preceded by other cue phrases, and either was deaccented or bore a **L*** pitch accent. Sentential uses also fit one of two prosodic models: in both, they were part of a larger intermediate phrase. In one model, they were

Table 13: Discourse/Sentential Models Using Part-of-Speech and Orthography.

Part-of-Speech	Model	N	Correct	
			Number	Percent
article	n=Sentential	6	6	100.0
coordinating conjunction	p=Discourse; n=Sentential	376	284	75.5
cardinal numeral	n=Sentential	21	21	100.0
subordinating conjunction	p=Discourse; n=Sentential	99	83	83.8
preposition	n=Sentential	3	3	100.0
adjective	n=Sentential	13	12	92.3
singular or mass noun	p/s=Discourse; n=Sentential	15	11	73.3
singular proper noun	p/b=Discourse; n=Sentential	6	5	83.3
intensifier	p=Discourse; n=Sentential	10	9	90.0
adverb	p/b=Discourse; s/n=Sentential	196	162	82.7
verb, base form	p=Discourse; n=Sentential	98	81	82.7
Total		843	677	80.3

Column 2 indicates the subdivisions of part-of-speech based on presence of adjacent orthography:
p – preceding
s – succeeding
b – both preceding and succeeding
n – no adjacent orthography

first-in-phrase and bore a **H*** or complex pitch accent — thus distinguishing them from discourse uses that were first-in-phrase. In the other, they were not first-in-phrase and bore any type pitch accent.

The association between discourse/sentential models and discourse/sentential judgments for this study, as for our previous studies of *now* and *well*, is significant at the .001 level. However, for the single-speaker, multi-cue phrase data in RJB86, our prosodic models successfully classified only 662 tokens (75.4%), a considerably smaller proportion than for the previous studies. We found one major reason for the poorer performance of our models on the multi-cue phrase data. A large percentage of the tokens that do not fit our prosodic models were coordinate conjunctions. When these are removed from our sample, our prosodic models correctly classify 442 tokens (85.3% of the data). It is also worth noting that coordinate conjunctions were among the most difficult cue phrases to classify as discourse or sentential.

To improve our notion of the factors that distinguish discourse from sentential uses, we made a more general examination of the set of items that we were unable to classify. In addition to the finding that conjunctions were difficult to classify (61 tokens, representing 81.3% of the tokens in RJB86 that we were unable to agree on a classification for), we also found that certain prosodic configurations appeared to make tokens more or less difficult to classify. Of the 75 unclassified tokens for RJB86, 55 (73.3%) were tokens of Discourse Model B or Sentential Model A. Recall that Discourse Model B identifies items that are first-in-phrase and are deaccented or bear a **L*** pitch accent; Sentential Model A identifies items that are also first-in-phrase but bear a **H*** or complex pitch accent. Discourse Model A, items that are alone in intermediate phrase, and Sentential Model B, items that are not first-in-phrase, appear easier to classify. Thus, it appears that prosodic configurations that are distinguished solely by differences in pitch accent, rather than upon differences in phrasing and position within a phrase, may be less useful indicators of the discourse/sentential distinction.

Furthermore, we found that orthographic cues (from transcription) successfully disambiguate between discourse and sentential usage in 675 cases (80.1% of the 843). Part-of-speech was less successful in distinguishing discourse from sentential use, disambiguating only 561 cases in the study (63.9% of 878). Using both orthography and part-of-speech for predicting the discourse/sentential distinction in our corpus was nearly equivalent to using orthography alone, predicting 677 (80.3% of 843) cases correctly. The relationship between the orthography of transcription and the orthography of written text will be an important determinant of whether orthography alone can be used for prediction in text-to-speech applications; if the latter is less useful, part-of-speech may provide additional power.

6 Discussion

In this paper, we have examined the problem of disambiguating cue phrases in both text and speech. We have presented results of several analyses of cue phrase usage in corpora of recorded, transcribed speech, in which we examined a number of text-based and prosodic features to find which best predicted a discourse/sentential distinction. Based on these studies, we have proposed an intonational model for cue phrase disambiguation in speech, based on intonational phrasing and pitch accent, and a model for cue phrase disambiguation in text, based on orthographic indicators and part-of-speech information.

Work on the meanings associated with particular intonational features, such as phrasing and pitch accent type, provides an explanation for the different prosodic configurations associated with discourse and sentential uses of cue phrases. As we have demonstrated above, discourse uses of cue phrases fit one of two models. In one model, Discourse Model A, discourses uses are set apart as separate intermediate phrases. Recall from

Section 3 that intonational phrasing can serve to divide speech into units of information, for purposes such as scope disambiguation. So, a broader discourse scope for a cue phrase may be signalled by setting it apart from other items which it might potentially modify if interpreted more narrowly. That is, in an utterance like *Now let's talk about cue phrases*, *now* may be more likely to be interpreted in its discourse sense if it is physically set apart from the verb it might otherwise modify in its sentential guise.

We have also seen that a discourse cue phrase may be part of a larger intermediate phrase and deaccented or given a **L*** pitch accent — Discourse Model B. While the absence of a pitch accent generally tends to convey that an item represents old information or is inferrable in the discourse, deaccenting is also frequently associated with function words — prepositions, pronouns, and articles. Cue phrases in the deaccented subset of Discourse Model B may, like function words, be seen as conveying structural information, rather than contributing to the semantic content of an utterance. The alternative version of Discourse Model B, in which a cue phrase that is part of a larger phrase receives a **L*** pitch accent, might be understood in terms of the interpretation proposed by Pierrehumbert and Hirschberg (1990) for the **L*** accent. In this account, the **L*** accent is analyzed as conveying that an item is salient in the discourse, but for some reason should not be added to speaker and hearer's mutual belief space. This subset of Discourse Model B cue phrases may thus be analyzed as conveying salient information about the discourse, but not adding to the semantic content of speaker and hearer's beliefs.

The text-based and prosodic models of cue phrases we have proposed from our studies of particular cue phrases spoken by multiple speakers, and of multiple cue phrases spoken by a single speaker, have both practical and theoretical import. From a practical point of view, the construction of both text-based and prosodic models permit improvement in the generation of synthetic speech from unrestricted text. From our text based model, we know *when* to convey a discourse or a sentential use of a given cue phrase. From our prosodic model, we know *how* to convey such a distinction. These distinctions have in fact been implemented in a new version of the Bell Labs Text-to-Speech System (Sproat, Hirschberg, and Yarowsky, 1992). From a theoretical point of view, our findings demonstrate the feasibility of cue phrase disambiguation in both text and speech and provide a model for how that disambiguation might be accomplished. These results strengthen the claim that the discourse structures crucial to computational models of interaction, in this case, certain lexical indicators of discourse structure, can indeed be identified.

Appendix

Table 14 summarizes the proposed meanings of items classed as cue words in six computational and linguistic treatments. Note that, we have omitted Cohen's discussion of Quirk's attitudinal expressions. Under 'Grosz/Sidner '86', we use **push**, **pop to** and **complete** to denote their attentional changes and the abbreviations "sat-pre" and "new-dom" for **satisfaction-precedes** and **new dominance**, respectively. Under 'Schiffrin '87', we use "marker" if the meaning of the discourse usage is illustrated via example, but is not discussed in detail. Under 'Warner '85', we use "conjunction" to denote his **simple conjunction** and "adversative" to denote his **adversative conjunction**.

Acknowledgments
We thank Ron Brachman for providing one of our corpora and Jan van Santen for helpful comments on this work. This work was partially supported by DARPA under contract N00039-84-C-0165.

References

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.

Beckman, Mary and Janet Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook*, 3:15–70.

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

Brazil, D., M. Coulthard, and C. Johns. 1980. *Discourse Intonation and Language Teaching*. Longman, London.

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

Church, K. W. 1988. A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the Second Conference on Applied Natural Language Processing*, pages 136–143, Austin. Association for Computational Linguistics.

Cohen, Robin. 1984. A computational theory of the function of clue words in argument understanding. In *Proceedings of Coling84*, pages 251–255, Stanford. International Conference on Computational Linguistics.

Francis, W. Nelson and Henry Kucera. 1982. *Frequency Analysis of English Usage*. Houghton Mifflin, Boston.

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.

Halliday, M. A. K. and Ruquaiya Hassan. 1976. *Cohesion in English*. Longman.

Hirschberg, J. 1991. *A Theory of Scalar Implicature*. Garland Publishing, Inc., New York.

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

Hirschberg, J. and G. 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.

Table 14: Suggested Meanings of Cue Phrases.

Cue Word	Cohen '84	Grosz/ Sidner '86	Reichman '85	Schiffrin '87	Warner '85	Zukerman/ Pearl '86
accordingly	inference					
again	parallel					
alright				marker		
also	parallel				conjunction	additive
alternately	reformulation					
alternatively						additive
although					adversative	adversative
altogether	summary					
and	parallel	push; new dom		addition; continuation; repair	conjunction	additive
anyway		pop to	return	response	hedge	
because			support	repair; resultive	causation	
boy				repair		
but	contrast	push	direct challenge	adversative; contrast; interruption; repair	adversative	adversative
consequently	inference				causation	
converseley	contrast					
equally	parallel					
finally	parallel	sat-pre; new dom				temporal
fine		complete				
first	parallel	sat-pre; new dom				
further	parallel					
furthermore	parallel	sat-pre; new dom				
gee				marker		
hence	inference					
hey				marker		
hopefully						causal
however	contrast			contrast	adversative	adversative
incidentally		digression	interruption			focal
indeed						additive
last	parallel					
like			support	comparison; repair; restriction	example; comparison	
likewise	parallel					
listen			prior logical abstraction	marker		
look				renewed initiative		
meanwhile	contrast					
moreover	parallel	sat-pre; new dom				
namely	reformulation					categorical
next	parallel	push				temporal
nevertheless						adversative

Table 14: Suggested Meanings of Cue Phrases, continued.

Cue Word	Cohen '84	Grosz/ Sidner '86	Reichman '85	Schiffrin '87	Warner '85	Zukerman/ Pearl '86
nonetheless	contrast					
nor					conjunction	
now		push	further development	progression; prominence; repair		
oh				repair		
ok		complete		marker		
only					adversative	
or				generalizer	alternation	additive
otherwise	contrast				conditional	
overall	summary					
say				marker		
second	parallel	sat-pre; new dom				
see				repair		
similarly	parallel					
still					adversative	
so	contrast		restatement; conclusion	development; repair; response; resultive	causation	causal
then	parallel			response		causal; temporal
therefore	inference; summary	new dom			causation	causal
though	contrast			response	adversative	
thus	summary					
too	parallel				conjunction	
unless					conditional	
well				repair; response		
where					example	
whereas					adversative; causation	
why				marker		
yet	contrast				adversative	

Hirschberg, Julia and Diane Litman. 1987. Now let's talk about *now*: Identifying cue phrases intonationally. In *Proceedings of the 25th Annual Meeting*, pages 163–171, Stanford University. Association for Computational Linguistics.

Hobbs, J. 1979. Coherence and coreference. *Cognitive Science*, 3(1):67–90.

Hockey, Beth Ann. 1991. Personal communication, July.

Litman, Diane and Julia Hirschberg. 1990. Disambiguating cue phrases in text and speech. In *Papers Presented to the 13th International Conference on Computational Linguistics*, pages 251–256, Helsinki, August. International Conference on Computational Linguistics.

Litman, Diane J. and James F. Allen. 1987. A plan recogntion model for subdialogues in conversation. *Cognitive Science*, 11:163–200.

Mann, W. C. and S. A. Thompson. 1983. Relational propositions in discourse. Technical Report ISI/RR-83-115, ISI/USC, November.

Olive, J. P. and M. Y. Liberman. 1985. Text to speech – an overview. *Journal of the Acoustic Society of America, Suppl. 1*, 78(Fall):s6.

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.

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.

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

Pollack, M. E., J. Hirschberg, and B. Webber. 1982. User participation in the reasoning processes of expert systems. Technical Report MS-CIS-82-9, University of Pennsylvania, July. A shorter version appears in the AAAI Proceedings, 1982.

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.

Quirk, R. 1972. *A Grammar of Contemporary English*. Longman, London.

Reichman, R. 1985. *Getting Computers to Talk Like You and Me: Discourse Context, Focus, and Semantics*. Bradford. The Massachusetts Institute of Technology, Cambridge.

Schegloff, E. A. 1979. The relevance of repair to syntax-for-conversation. In T. Givon, editor, *Syntax and Semantics*, volume 12. The Academic Press, New York, pages 261–288.

Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge University Press, Cambridge UK.

Schourup, Lawrence. 1985. *Common Discourse Particles in English Conversation*. Garland Publishing, Inc., New York.

Sidner, C. L. 1985. Plan parsing for intended response recognition in discourse. *Computational Intelligence*, 1(1):1–10.

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. 1992a. TOBI: A standard scheme for labeling prosody. In *Proceedings of the Second International Conference on Spoken Language Processing*, Banff, October. ICSLP.

Silverman, Kim E. A., Eleonora Blaauw, Judith Spitz, and John F. Pitrelli. 1992b. Towards using prosody in speech recognition/understanding systems: Differences between read and spontaneous speech. In *Proceedings of the Speech and Natural Language Workshop*, pages 435–440, Harriman NY, February. DARPA, Morgan Kaufmann.

Sproat, R., J. Hirschberg, and D. Yarowsky. 1992. A corpus-based synthesizer. In *Proceedings of the International Conference on Spoken Language Processing*, Banff, October. ICSLP.

Swora, Maria G. and Mary E. Beckman. 1991. The intonation of cue words in task-oriented dialogues. Presented at the LSA Annual Meeting, Chicago, January.

Talkin, David. 1989. Looking at speech. *Speech Technology*, 4:74–77, April-May.

Warner, Richard G. 1985. *Discourse Connectives in English.* Garland Publishing, Inc., New York.

Zuckerman, Ingrid and Judea Pearl. 1986. Comprehension-driven generation of meta-technical utterances in math tutoring. In *Proceedings of the Fifth National Conference*, pages 606–611, Philadelphia. AAAI.