Intonational overload: Uses of the Downstepped (H* !H* L- L%) contour in read and spontaneous speech

Julia Hirschberg, Agustín Gravano, Ani Nenkova, Elisa Sneed and Gregory Ward

Abstract

Intonational contours are *overloaded*, conveying different meanings in different contexts. In this paper we examine two potential uses of the *downstepped* contours in Standard American English, in the Boston Directions Corpus of read and spontaneous speech. We investigate speakers' use of these contours in conveying discourse topic structure and in signaling *given* vs. *new* information and discuss the possible relationship between these two functions.

1. Introduction

It is well known that a single intonational contour may convey different meanings in different contexts: In Standard American English (SAE), for example, the simple declarative contour H* L- L% can be used felicitously over *wh*-questions as well as statements.[1] Similarly, the rise-fall-rise (L*+H L- H%) contour can convey either uncertainty or incredulity depending on the speaker's pitch range and voice quality (Sag and Liberman 1975, Hirschberg and Ward 1992, Nickerson and Chu-Carroll 1999). That is, intonational contours are *overloaded*. For the most part, however, it has proven difficult to find a single all-encompassing meaning for any given contour or to identify all and only the felicitous contexts for the appropriate use of that contour. In this paper we examine the so-called *downstepped* contours with respect to their function in signaling discourse information. We study them in a corpus of read and spontaneous monologues in a direction-giving domain in the Boston Directions Corpus (Nakatani, Grosz, and Hirschberg 1995; Hirschberg and Nakatani 1996).

While downstepped contours are widely used in SAE, the conditions under which they are likely to be produced have rarely been studied. In Pierrehumbert 1980's description, downstep in SAE may be triggered by

any *complex* pitch accent (Pierrehumbert 1980, Liberman and Pierrehumbert 1984). The most commonly produced examples in SAE are produced as sequences of H*+L accents, as represented in Pierrehumbert's (1980) model of SAE, and as H* !H* sequences as represented in the ToBI standard. These contours may end with a fall (H* !H* L- L% or a rise (H* !H* L-H%), and may be observed over full intonational phrases such as these or over intermediate phrases such as H* !H* L-. Downstepped contours may be triggered by other complex pitch accents, such as the L*+H L*+!H L- L% contour as well. We will refer to the set of all downstepped contours, whatever their accent type, as All-DS below; the downstepped contours containing only H* pitch accents (e.g. !H*) we will term DS contours, since specific predictions have been made both about the general class and the specific subtype. An example of the most common of these, H* !H* L-L%, is shown in Figure 1:

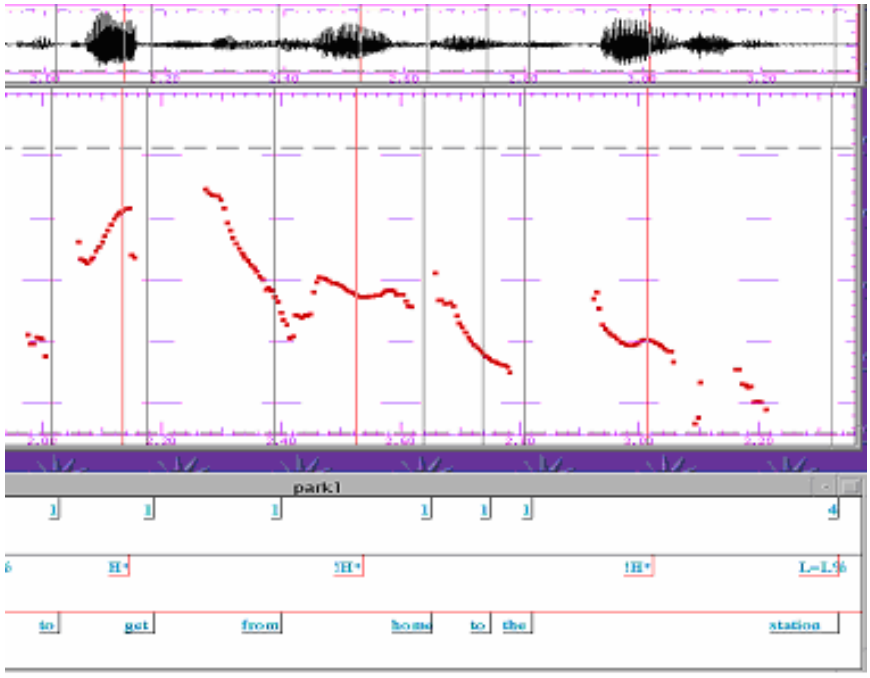

Figure 1. A H* !H* !H* L- L% Contour

The F0 of the H* !H* L-L% variant looks like a flight of steps. This intonational contour and the H* !H* L- contour, which represents a component intermediate phrase within the full intonational contour, appear

to be the most frequent contours in the !H* family. For example, in the AT&T Communicator Corpus of read speech (Hirschberg and Rambow 2001, Hastie et al. 2002), the H* !H* L- and H* !H* L- L% contours represent the most frequent pattern of the 2888 intermediate phrases in this 67-minute corpus, comprising about 40% (317/810) of all contours (Venditti 2002). They occur almost twice as often as the 'standard' declarative contours (H* L-L% and H* L-) in this corpus.

Despite their frequency of use, however, the circumstances under which downstepped contours are used, and the meanings associated with them by speakers and hearers, have not been systematically investigated, either in laboratory or in corpus-based studies. It has been speculated that downstepped contours mark discourse topic structure, occurring frequently in phrases which signal topic beginnings and endings (Pierrehumbert and Hirschberg 1990). It has also been proposed in that work that the interpretation of sequences of downstepped pitch accents of the DS type might be characterized as conveying that the hearer should be able to infer, from the beliefs the hearer and speaker share, the existence of discourse entities realized with such accents. A possibly related observation is that DS contours serve as an alternative to *deaccenting*, when information being expressed represents *given* information (Prince 1981, Prince 1992) in the discourse (Hirschberg and Pierrehumbert 1986). Such alternation has been observed in read speech collected by Dahan and colleagues (2002) as experimental materials for eye-tracking studies of the processing of information status. Ladd (1996) has further suggested that the downstepped contours may be used when speakers mention given information whose deaccenting would produce an undesirable alteration of the focus structure of the phrase. Finally, it has also been observed anecdotally that, when native speakers of SAE employ a DS contour, they convey a 'professorial', rather smug, didactic tone. A similar impression is reported for SAE speakers who interpret these contours in the speech of speakers of British Received Pronunciation (RP), for whom DS is the most common declarative pattern (Pierrehumbert and Hirschberg 1990).

In this paper we study the functions of the downstepped contours in the Boston Directions Corpus (BDC), a corpus of read and spontaneous speech collected and annotated for the study of intonational cues to discourse structure (Nakatani, Grosz, and Hirschberg 1995; Hirschberg and Nakatani 1996). In the work presented here, we examine several hypothesized functions for downstepped contours in the BDC: First, we examine the hypothesis that downstepped contours serve the discourse function of

introducing new topics or closing old topics. We investigate these hypotheses in read and spontaneous speech, and study these across different speakers. We then examine the hypothesis that DS contours may alternate with deaccenting to convey the givenness of information, where given information is defined as in (Prince 1992) with respect to either the hearer or the discourse. We will explore this hypothesis in terms of the use of deaccenting vs. production of a DS contour over NPs that are *Hearer old* or *Hearer inferable* or *Discourse old* in our corpus. We conclude with some preliminary attempts to identify how these and other utterance features account for the use of downstepped contours in this corpus.

2. The Boston Directions Corpus

The current investigation makes use of a corpus of spontaneous and read speech, the Boston Directions Corpus (BDC).[2] This corpus comprises elicited monologues produced by four non-professional speakers, three male and one female, who were given written instructions to perform a series of nine increasingly complex direction-giving tasks. Speakers first explained simple routes such as getting from one station to another on the subway, and progressed gradually to the most complex task of planning a round-trip journey from Harvard Square to several Boston tourist sights. Thus, the tasks were designed to require increasing levels of planning complexity. The speakers were provided with various maps, and could write notes to themselves as well as trace routes on the maps. For the duration of the experiment, the speakers were in face-to-face contact with a silent partner (a confederate) who traced on her map the routes described by the speakers. The speech was subsequently orthographically transcribed, with false starts and other speech errors repaired or omitted; subjects returned several weeks after their first recording to read aloud from transcriptions of their own directions. A total of 50 minutes of read speech and 66.6 minutes of spontaneous was collected, with speakers ranging from 7.9 to 17.9 minutes for the read tasks and 11.2 to 22.8 for spontaneous productions, with Speaker 3 producing the least speech and Speaker 2 the most in each case.

2.1. Prosodic analysis

The BDC was labeled for intonational features using the ToBI labeling scheme for SAE (Pitrelli et al. 1994, Beckman, Hirschberg, and Shattuck-Hufnagel 2004) from an F_0 contour calculated using Entropic's *get_f0* pitch tracker (Talkin 1989). 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 juncture 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 F_0 contour define intonational phrases, following Pierrehumbert's (1980) scheme for describing SAE (with some modifications) and a *miscellaneous tier*, in which phenomena such as disfluencies may be optionally marked. [3] Of primary interest for this study is our use of the tones and break index tiers to identify ToBI level 3 and 4 phrases and the pitch accents, phrase accents, and boundary tones included in them. Level 4 (corresponding to Pierrehumbert's intonational 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-). The downstepped contours we are examining in this paper, for example, end in a low phrase accent (L-), a low phrase accent and low boundary tone (L- L%) or a low phrase accent and high boundary tone (L- H%).

Pitch accents render words intonationally prominent and are realized by increased F_0 height, loudness, and duration of accented syllables. Any word may be accented or deaccented (Ladd 1979) and, if accented, may bear different tones, or different types of prominence, with respect to other words. Five types of pitch accent are distinguished in the ToBI system for American English: two simple accents H* and L*, and three complex ones, L*+H, L+H*, and H+!H*. The asterisk indicates which tone of the accent is aligned with the stressable syllable of the lexical item bearing the accent. Pierrehumbert's complex H*+L accent is included in ToBI's H* category, and is distinguished contextually from H* by the presence of a following downstepped tone (!H*). Downstepped accents follow a complex pitch accent and occur in a pitch range that is compressed in comparison to a non-downstepped accent. Downstepped accents are indicated by the '!' diacritic in the accent label. So, the downstepped accents we are examining

here can be represented equivalently as a sequence of H*+L H* in Pierrehumbert's representation and as H* !H* in the ToBI system.

2.2. Discourse segmentation

The BDC corpus was also segmented according to the Grosz and Sidner 1986 (G&S) theory of discourse structure, which provides a theoretical basis for segmenting discourses into its component parts. The G&S model defines discourse structure as consisting of a series of *discourse segments*, defined in terms of a speaker's underlying intentions in producing each segment; for each discourse segment there is a corresponding *discourse segment purpose* (DSP). These segments are related to one another in terms of the relationship of their DSPs, which may be one of the following types: 1) a DSP A *satisfaction-precedes* a DSP B if A must first be achieved in order for DSP B to be successful; and 2) a DSP A *dominates* a DSP B if fulfilling B partly fulfills A. Thus, segments may be related to one another as siblings or as children, depending on the relationships of their DSPs. The segments and the embedding relationships between them form G&S's *linguistic structure*. The embedding relationships reflect changes in the *attentional state*, the dynamic record of the entities and attributes that are salient during a particular part of the discourse. Changes in linguistic structure, and hence attentional state, depend on the *intentional structure* of the discourse, which comprises the DSPs underlying the discourse and relations among DSPs. Each discourse is posited to reflect a single *discourse purpose*.

The discourse structure of the BDC was annotated according to this theory by two groups of annotators, 'expert' and 'naive'. The expert labelers were all knowledgeable about the G&S theory, and were given minimal instructions; they annotated only one speaker's data. The naive group consisted of nine Harvard undergraduates, with no previous knowledge of G&S theory. They were provided with a labeling manual which gave an overview of the theory with detailed examples of annotations (Nakatani, Grosz, and Hirschberg 1995). [4] These labelers labeled four speakers' productions. In both expert and naive labeling, three annotators labeled each speaker task, with no labeler labeling both read and spontaneous versions of any speaker task. Both sets of labelers could listen to the original speech as well as read the transcription while labeling. [5] An example of one labeler's segmentation of a short speaker task is shown

below, where indentation is used to indicate the hierarchical relationships between segments:

D1 DSP1 : GET ON AT HS FIRST
get on the Harvard Square T stop

D2 DSP2 MORE DETAIL ON GETTING ON
uh get on the har(vard)- uh get on the subway at the Harvard Square T stop

D3 DSP3: BUY TOKEN
and purchase a token

D3.5 DSP3.5 GO INBOUND ON RED LINE
and go on the Red Line
inbound

D4 DSP4: TO PARK STREET
go from
Harvard Square
to the Park Street Station

D5 DSP5: GET OFF SUBWAY
then
get off the subway
(the Red Line subway)

D6 DSP6: GET ON GREENLINE
and get on the Green Line subway

D6.5 DSP6.5 GO TO COPLEY
going
to Copley Station

D7 DSP7 IDENTIFYING CARS AS EQUAL TO COPLEY
any of the um
any of the different Green Line
cars
will take you to Copley Station

D8 DSP8: BOARD A CAR
so simply board one

D9 DSP9: take car to B &A
and take it through Boylston
and Arlington

D10 DSP10: END AT CS
and then on to Copley Station

For our study, we used only the naive labelers' annotations, so that we could include more of the data in the study. Since the interlabeler reliability of these labelers' judgments was less than for our expert labelers (e.g. .58 agreement for judgments of segment beginnings for the spontaneous productions across all tasks and speakers and only .45 agreement for read productions), we considered only majority decisions of the annotators — i.e., decisions about segment boundary beginnings or endings for which at least two of the three annotators' decisions were in agreement. While labelers annotated each task hierarchically, we consider here only majority agreement on whether or not an intermediate phrase constituted a segment-beginning (SBEG) or a segment-final (SF) phrase.

2.3. Further annotation

The spontaneous speech data used for the current study has been further hand-labeled for given/new information status, using Prince's (1992) distinctions of hearer-old/hearer-new and discourse-old/discourse-new status for each simple NP in the corpus. In addition, each *discourse element*, or description of a discourse entity or entity feature within an NP, was coded for its discourse status (i.e. old or new).

For Prince (1992), Discourse-Old (DO) information is that which has been explicitly or implicitly evoked in the prior discourse, whereas Discourse-New (DN) information is that which has not been previously evoked. Hearer-Old (HO) information, regardless of whether it has been evoked in the current discourse, is assumed to be known to the hearer, while Hearer-New (HN) information is assumed to be new to the hearer. Lastly, Hearer-Inferable (HI) information is that which is not expected to be known to the hearer, but which the speaker can infer based on other DO elements that trigger its existence.

For each concrete noun in the corpus, two coders assessed its information status along these two dimensions for hearer and discourse status. (So as to prevent possible bias, nouns were coded for their information status without any prosodic information available to the coders.) Two sets of transcripts were coded. The first was coded for hearer and discourse status by labeling the information status of the entire NP constituent. The other set was coded only for discourse status, by coding the information status of each content word within an NP. Our motivation for choosing this coding schema was that we wanted to represent the

discourse status of each NP, whether or not it was contained within a larger NP.

For hearer status, we considered each maximal NP and tried to determine whether the entity corresponding to that NP constituted familiar information for the hearer; that is, whether or not the speaker had sufficient evidence for the hearer to know about (or be familiar with) that entity. For discourse status, each noun (and its pre-nominal modifiers) was coded for whether it had appeared in the previous discourse. The nature of the monologues in the BDC informed our use of coding conventions in a number of ways. For instance, the first mention of "T" stations, as well as transit lines, were all coded as HO (but DN) because both speakers/hearers could see the stations on the system map displayed in front of them. Moreover, as typical Bostonians/Harvard students, the speakers/hearers were already, presumably, familiar with the lines/stops. In order to determine the hearer status of the various streets mentioned in the corpus, we enlisted the help of an undergraduate at Harvard, who asked her friends whether or not they knew of the streets in question. Those streets that were deemed by them to be familiar were coded as HO. Discourse entities that did not appear to be discourse initial (e.g. *Newbury Comics* in: "We have just left Newbury Comics"), were coded as DN, with their Hearer status determined either by presumed world knowledge (e.g. Logan Airport = HO) or by our student informants (e.g. Newbury Comics = HO).

3. The Downstepped contour and discourse structure

The first hypothesis we explore is the proposal that downstepped contours are an important cue to the intentional structure of a discourse, signaling when new discourse segments begin or when they end. To investigate this hypothesis, we examined majority segmentation decisions of naive labelers on all four speakers' spontaneous and read productions. Specifically, we wanted to know: How often do the speakers use downstepped contours to mark topic beginnings and topic endings? Does this usage differ in read vs. spontaneous speech? Are there speaker differences in the use of downstepped contours?

3.1. Read and spontaneous productions across all speakers

There are 3183 intermediate (ToBI level 3) phrases in the spontaneous productions of the four speakers, 552 for Speaker 1, 1135 for Speaker 2, 567 for Speaker 3, and 929 for Speaker 4. Read speech was shorter for all speakers in time as well as in number of intermediate phrases: In the corresponding read productions of the speakers, prosodic labelers found a total of 2153 intermediate phrases: 495 for Speaker 1, 752 for Speaker 2, 361 for Speaker 3, and 545 for Speaker 4.

Looking at patterns of use of downstepped contours across all four speakers, we note first that, in general, All-DS (including all types of downstep) contours appear more frequently in read speech than in spontaneous, with approximately half (49%) of read phrases and just over one third (37%) of spontaneous productions in our corpus produced with downstep. The DS contour (only those with H* !H* pitch accents) comprises about 40% of All-DS contours in each speaking condition and is distributed similarly, forming 21% of read productions and 15% of spontaneous.

When we look at how the downstepped contours are distributed with respect to discourse function as compared to the contours characterized by simple H* pitch accents (including H* L- L%, H* L- H%, and H* L-), we see that DS contours pattern much like the simple H*, with 29% of each contour group occurring with segment beginning (SBEG) phrases in read speech and 18% each in spontaneous speech. However, the similarlity does not hold for segment final (SF) phrases, as seen in Figure 2.

While 36% of H* contours appear in SF positions in read speech, 43% of read DS contours occur in SF phrases; and the difference in spontaneous speech is more marked, with 28% of H* contours appearing in this position, while fully 40% of DS contours do. So, while the proposal that DS contours mark segment beginnings and endings is not borne out in these data for either read or spontaneous speech — they appear similar to simple H* contours in this regard — there does seem to be a predilection for using DS contours over H* contours to signal segment finality (for read: chisq=31.76, df=1, p=0; for spontaneous: chisq=18.00, df=1, p=0). However, when we include all downstepped contours (All-DS) in the analysis, we see that a much larger proportion of these occur in SBEG position, with 34% in read and 26% in spontaneous speech. In fact, the Other DS contours alone show a greater propensity for SBEG position than the DS contours: 38% of them occur in SBEG position in read speech and

31% in spontaneous; the difference between DS and Other DS

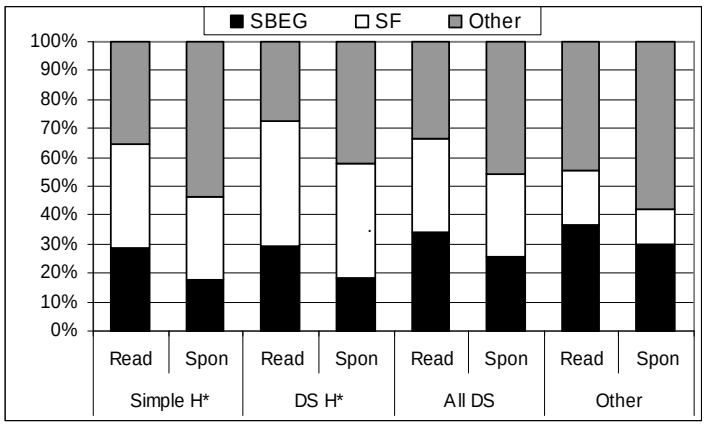

Figure 2. Distribution of contours by segment position for read and spontaneous speech

contours is significant (read: chisq=4.59, df.=1; p=.032; spontaneous: chisq=26.51, df=1, p=0). However, for All-DS contours in general, there is less indication of an association with segment finality. The proportion of All-DS contours used in SF position is roughly the same as the proportion appearing in SBEG position — 33% for read speech and 28% for spontaneous. The Other DS contours are employed in SF position 25% of the time in read speech and only 20% in spontaneous.

Thus, while we have some support for the notion that downstep signals discourse segment beginning, it appears to be the non-DS contours and not the DS contours which are used with some frequency in this position. However, a larger proportion of DS contours *do* appear to be used with segment-final phrases than any other contour type, including the simple H* contours and the Other DS contours. This difference is particularly notable in spontaneous speech.

Turning now to the related question of how often SBEG or SF phrases are uttered with some form of downstep vs. other contours, we see clearly in Figure 3 that downstepped contours (All-DS) do dominate the production of both SBEG and SF phrases in read speech but are less important in SBEG productions in spontaneous speech.

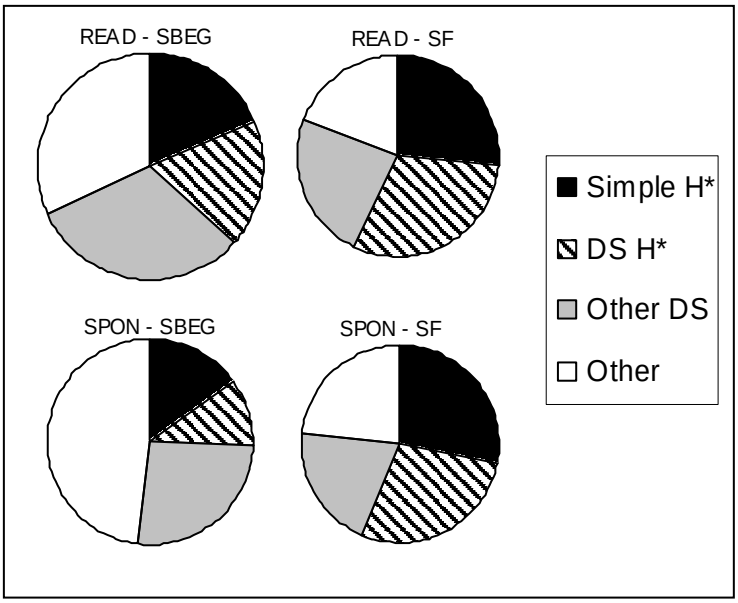

Figure 3. Proportion of segment beginnings and endings by contour, read and spontaneous speech

In read speech, simple H* contours and DS contours appear in similar numbers for SBEG phrases (19% of SBEG phrases are uttered with H* contours and 18% with DS) and SF phrases (27% and 31% for H* and DS, respectively. However, All-DS contours constitute fully half of read SBEG phrases (50%) and over half of read SF phrases (54%). While DS contours account for only 11% of SBEG phrases in spontaneous speech and 28% of SF phrases, All-DS contours make up fully 37% of SBEG phrases and 49% of SF phrases in spontaneous productions. So, again, while the particular DS contours hypothesized in the literature as signaling discourse structure do appear to play a larger role at least in segment final phrases, it is the more general class of all downstepped contours (All-DS) that figures most prominently in both segment-beginning and segment-final productions. And for all but spontaneous topic-beginning phrases, downstepped contours form the largest category of phrases in these positions.

The general role of downstep in signaling segment beginnings and endings thus has considerable support from our data, if the details are somewhat different than those suggested in the literature. While the DS contours (realized with H* !H* pitch accents) do appear important in themselves in signaling segment finality, the larger class of All-DS

contours seems to be even more important, particularly in signaling segment beginnings, where the DS contours are represented even less well than the simple H* contours.

3.2. Speaker variability

In the context of these overall findings about the general usage of the downstepped contours in the BDC corpus, we next investigate whether there are speaker differences in the production of DS and All-DS contours, and in SBEG and SF discourse positions in particular. We should first note that there are considerable differences in overall contour use among our four speakers. Table 1 shows the distribution of downstepped (DS and All-DS) and simple H* contours by speaker, for both read and spontaneous productions. Percentages shown indicate the proportion of phrases in each contour category.

For all four speakers, downstepped contours are the most frequently used contour type in read speech, compared with all other contours; All-DS contours represent 42-54% of each speaker's productions. However, speakers vary more widely in their use of the particular class of DS contours, where usage ranges from 15-34%. Simple H* contours, generally considered the most frequent contour type in SAE, make up roughly similar percentages of three speakers' intonational repertories, with only Speaker 3 employing a markedly higher proportion of simple H* contours compared to DS; this speaker employs Other DS contours twice as often as DS. So, in read speech, downstepped contours represent the majority contour type for our speakers. In spontaneous speech, downstepped contours form a lesser proportion of all speakers' productions, ranging from 25-43%. However, simple H* contours do not dominate for any speaker, even in this genre, forming only 15-31% of productions. DS contours do appear much less frequently in spontaneous speech than simple H* contours, except for Speaker 2, for whom they appear in almost identical numbers. So, in general, downstep appears to be more common in read speech, although still well represented in spontaneous speech, particularly when compared to simple H* contours. The DS contours are somewhat better represented in read speech than in spontaneous, but there is clearly considerable variation among speakers in both conditions.

Table 1. Distribution of contours by speaker

	Contour	READ		SPON		TOTAL
		N	%	N	%	
Speaker 1	Simple H*	158	32%	169	31%	327
	DS H*	166	34%	110	20%	276
	All-DS	265	54%	237	43%	502
	Total	495		552		1047
Speaker 2	Simple H*	124	16%	248	22%	372
	DS H*	148	20%	242	21%	390
	All-DS	354	47%	431	38%	785
	Total	752		1135		1887
Speaker 3	Simple H*	101	28%	85	15%	186
	DS H*	54	15%	27	5%	81
	All-DS	153	42%	143	25%	296
	Total	361		567		928
Speaker 4	Simple H*	83	15%	183	20%	266
	DS H*	83	15%	109	12%	192
	All-DS	282	52%	369	40%	651
	Total	545		929		1474

Figures 4a and 4b speak to the question "How does a particular speaker employ a particular contour in the various discourse positions (SBEG, SF and Other)?" for read and spontaneous speech. These graphs show that different speakers use DS contours in conveying discourse segmentation very similarly, except with respect to SBEG phrases in read speech. All speakers use around 40-50% of their DS contours over SF phrases, whether in read or spontaneous speech. In spontaneous speech, the use of DS contours ranges from 15-25% for SBEG phrases. However, in read speech, the use of DS contours over these SBEG phrases constitutes from 16%-42% of total DS use, depending on speaker, with Speaker 1 employing DS 42% of the time over SBEG phrases. Since Speaker 1 also has the largest proportion of DS contours in his productions, this additional use may account for this. It appears that the use of DS contours to signal segment beginnings may be idiosyncratic for certain speakers. Looking at the more general class of All-DS contours, we see that, in read speech and for all speakers, when we include other downstepped contours, the proportion of downstep used in SBEG and in SF phrases increases — for Speaker 2,

quite dramatically, from 16% of DS contours used in SBEG phrases to 28% of All-DS. In spontaneous speech, All-DS contours affect primarily the proportion of downstepped contours used in SBEG phrases for two speakers, with two others showing an effect in both SBEG and SF similar to what we see in read speech.

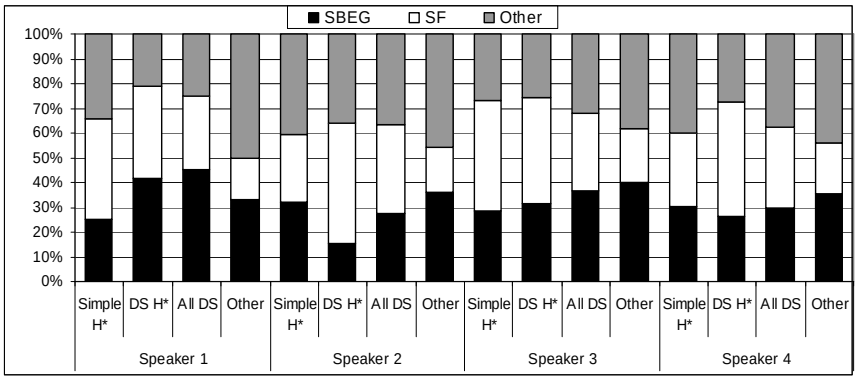

Figure 4a. Proportion of contours by segment position and speaker, read speech

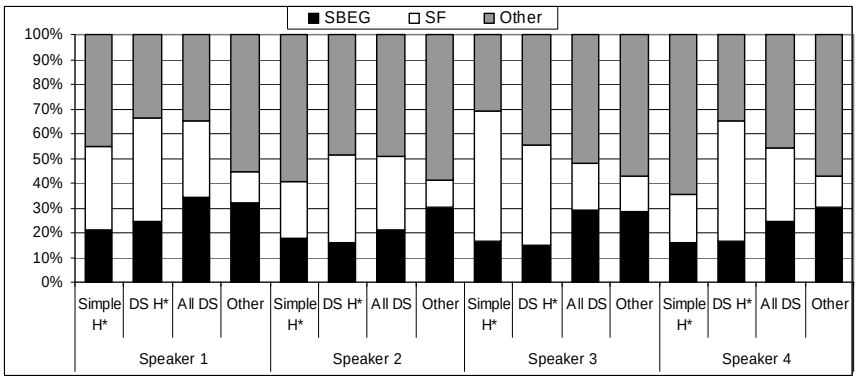

Figure 4b. Proportion of contours by segment position and speaker, spontaneous speech.

We now turn to an examination of individual differences in the proportion of DS contours that speakers use to signal SBEG or SF, compared to alternate contours: How important are DS contours in the overall production of SBEG and SF phrases? Figure 5 shows the proportion of SBEG and SF phrases uttered with simple H*, DS, Other DS contours and all other contours, by speaker.

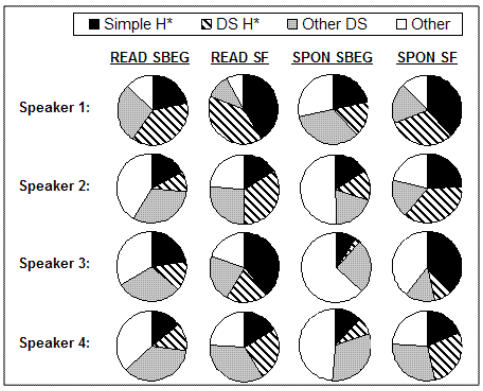

Figure 5. Proportion of contours in SBEG and SF utterances by speaker, read and spontaneous speech.

Figure 5 further confirms the individual differences observed earlier. Looking first at DS contours alone, we find in read speech that there is a considerable difference between Speaker 1 and the other speakers in the proportion of SBEG phrases uttered with DS; this difference is enhanced when we look at All-DS contours. The same difference is found in SF phrases, where 51% of this speaker's SF phrases are uttered with some form of downstepped contour. In spontaneous speech, we also see large differences among speakers for both SBEG and SF phrases. Note, in particular, the very small (3%) proportion of SBEG and SF phrases uttered with DS by Speaker 3, compared to the other speakers, and the small proportion of SF phrases this speaker utters with All-DS compared to the other speakers. Speaker 3 is also the only speaker to use a considerably smaller proportion of downstepped contours over SF phrases in spontaneous speech compared to read speech, although all speakers' SBEG phrases are more often uttered with All-DS in read as compared to spontaneous speech.

3.3. Discussion

In general, the hypothesis that speakers employ a particular type of downstepped contour, which corresponds to those downstepped contours with !H* pitch accents (DS), to signal discourse topic beginnings is not supported by our data, while the hypothesis that they do use this contour to

signal topic finality is. However, if we broaden this hypothesis to include all downstepped contours (ALL-DS), the role of downstep in signaling segment beginnings has more support. These Other DS contours (e.g. L*+H !H*) appear more important in fact in indicating SBEG phrases than SF.

We also find significant differences between genres in the use of downstepped contours in read vs. spontaneous speech: ALL-DS contours appear more frequently in read than in spontaneous speech — in fact, they are the majority contour type for read speech — and DS contours follow the pattern of their parent type. There are individual speaker differences as well as differences in use by genre, however. Individual speakers vary in their overall use of DS contours for any type of phrase and any condition rather considerably.

4. The Downstepped contour and given/new status

We next investigate whether downstepped contours — DS or the broader All-DS category — are associated with differences in information status. In particular, is a downstepped contour used over given information, alternating with a deaccenting strategy? If so, when do speakers choose one strategy over another? To answer these questions, we looked at information status at the NP level, both in terms of discourse and hearer-centered given/new status, and at the word level, in terms of discourse given/new status for individual lexical items. Note then that any NP has both a hearer-based and a discourse-based given/new status.

Our corpus is smaller for this aspect of our study, since we have annotation of information status only for the spontaneous productions of the BDC corpus. Of the 1551 NPs in the spontaneous part of the BDC corpus, just under half (49%) are uttered with some type of downstepped contour (All-DS). Table 2 presents the distribution of these and other contours by information status for all NPs. In Tables 2 and 3, HG denotes 'Hearer-Given' and HN 'Hearer-New'. Similarly for DG and DN. HI denotes 'Hearer-Inferable', where inferable discourse entities are defined as in (Prince 1992) to be those whose existence can be inferred from the existence of other entities already evoked in the discourse. The 'All Deacc' row represents NPs in which all discourse elements (content words) are deaccented; the table shows that this is a very small category.

Table 2. Distribution of NPs by information status and contour type

Contour	HG		HI		HN		DG		DN	
DS	416	41%	200	49%	58	45%	261	44%	413	44%
Other DS	48	5%	25	6%	12	9%	32	5%	53	6%
All-DS	464	46%	225	55%	70	54%	293	49%	466	49%
All Deacc	5	.05%	6	2%	3	2%	46	8%	15	2%
Other	540	54%	175	43%	57	44%	257	43%	469	49%
Total	1009		406		130		596		950	

Note that DS contours constitute a large fraction of all status categories, both given and new, at both the Hearer and Discourse levels of analysis. This fairly even distribution of DS in both given and new NPs is a strong confirmation that DS is at least multifunctional. Even when we consider all downstepped contours (All-DS), the distributions remain almost equal across given and new NPs both at the Hearer and the Discourse level; most contours uttered over NPs are, in fact, of the DS variety. When we compare the distribution of contours within each given/new category — e.g. what proportion of HG or DN NPs are uttered with downstep vs. other contours — we see only that HI NPs are more likely to be downstepped than not. This finding would support the hypothesis in Pierrehumbert and Hirschberg (1990) that DS contours are used to convey that discourse entities should be inferable from speaker and hearer's shared beliefs.

It is interesting to note that, in support of Ladd (1996), given NPs are rarely fully deaccented. Table 3 presents contours for NPs in which all individual discourse elements are labeled as given, and the proportion of these that are also fully deaccented. Not surprisingly, there are few Hearer- or Discourse-New NPs in this table.

Table 3. Contours of NPs for which all elements are Given

Contour	HG		HI		HN		DG		DN	
DS	260	45%	38	54%	3	33%	251	45%	50	52%
Other DS	28	5%	2	3%	2	22%	28	5%	4	4%
All-DS	288	50%	40	57%	5	56%	279	50%	54	56%
All Deacc	45	8%	3	4%	0	0%	44	8%	4	4%
Other	244	42%	27	39%	4	44%	237	42%	38	40%
Total	577		70		9		560		96	

For both the Hearer-Given and Discourse-Given NPs, DS contours are represented in the majority of productions. However, while DS is the majority pattern for NPs with 'all given' elements and, in particular, is

much more likely to be employed than the complete deaccenting of all given elements in the NP, other patterns are represented almost as often as DS contours. If we again include other downstepped contours (All-DS), the predominance of downstep becomes stronger, rising to one half of Hearer-Given all-given NPs, 57% of Hearer-Inferables and 50% of Discourse-Given all-given NPs. However, about the same proportion of (Hearer and Discourse) New NPs are also uttered with All-DS. The only category of information status where DS and downstep in general clearly dominate, again, is the case of the Hearer-Inferables, in which DS contours represent more than half of all productions. However, this category itself is small. So, while we can conclude that DS contours are commonly used over given information, we have little evidence from this study that information status represents a major predictor of the use of DS, in and of itself, since they are equally likely to be used over New NPs.

4.1. Other Factors in Downstepped NPs

Given the lack of evidence that information status alone is a strong cue to the use of DS over NPs, but given also the frequency with which both Given and New NPs are uttered with DS, what additional factors might help us to understand when NPs are downstepped and when they are not? We have already seen that DS **does** have its use as an indicator of discourse structure and that there are differences also in the general proclivity of speakers to employ this contour. So, topic position and speaker identity might help to refine our general findings with respect to the production of NPs. However, since we have information status labels only for spontaneous productions, and since these showed a weaker correlation with topic structure than did read speech, it is not surprising that we find only suggestive rather than clear relationships between downstep and topic structure in our smaller corpus of spontaneous NPs than we did in the full corpus: Only when we include all downstepped contours (ALL-DS) do we find a significant correlation between downstepping and discourse structure position of NPs. Interestingly, this correlation is found only between All-DS and topic beginnings (SBEG) (chisq=6.31, df=1, $p<.01$). There is, however, a strong relationship between choice of DS contour to realize NPs and speaker identity (chisq=26.81, df=3, p=0). This distribution is shown in Table 4:

Table 4. Distribution of DS NPs by Speaker.

	Speaker 1	Speaker 2	Speaker 3	Speaker 4
DS	134(39.6%)	274(49.0%)	73(30.7%)	195(46.9%)
Non-DS	204	285	165	221
Total	338	559	238	416

While in all cases, speakers employ DS over NPs much more often than they do over contours in general (recalling distributions from Table 1) there are still clear differences between Speaker 3 and the other speakers.

We might also imagine that the length of an NP might play a role in whether or not it is downstepped, since a phrase must have at least two accented words in it in order to establish a downstepped contour. Indeed, a comparison of the length of NPs that are downstepped vs. those that are not shows a significant difference, although only for the broader family of All-DS contours: These contours are uttered over longer NPs than other contours ($t_{stat}=4.10$, $df=1549$, $p<.001$).

Putting some of these factors together gives us a more unified picture of what may account for the use of DS contours and downstepped contours in general (All-DS) over NPs. A logistic regression linear model analysis with DS (or no-DS) as the dependent binary variable and topic position of the NP (SBEG and SF), the length of the NP in words and in discourse entities, speaker identity, and the Hearer and Discourse status of the NP as independent variables, shows effects for several of these potential predictors (added sequentially, from first to last in the table). Table 5 presents these results; Table 6 presents a similar analysis for All-DS contours.

From Table 5 we see that Hearer-based given/new status and speaker identity are the only variables significantly associated with the prediction of DS contours in a phrase, although topic-initial (SBEG) position, Discourse-based given/new status, and number of discourse entities in the phrase tend to significance. We see no effect, notably, for segment-final (SF) position or for number of total words in the phrase. Contrast these findings, however, with our previous analysis of DS contours as a whole in the corpus in SBEG and SF position, where we found that DS contours in general appeared to mark SF but **not** SBEG phrases. For NPs, a high proportion (48%) of SBEG phrases are uttered with DS. There are also significant interactions between SF position and speaker, Hearer given/new status and number of words, and Discourse status and number of words.

Table 5. Linear Model Predicting DS contours over phrases

Predictor	**DF**	**Deviance**	**Residual DF**	**Residual Dev.**	**Pr(CHI)**
NULL			1550	2124.54	
SBEG	1	2.88	1549	2121.66	0.09
SF	1	0.09	1548	2121.58	0.77
H-STATUS	3	7.92	1545	2113.66	0.05
D-STATUS	2	5/10	1543	2108.56	0.08
SPEAKER	3	24.87	1540	2083.69	0.00
ENTITIES	1	2.56	1539	2081.13	0.11
WORDS	1	0.62	1538	2080.51	0.43
SF:SPEAKER	3	17.30	1524	2052.74	0.00
SF:ENTITIES	1	3.19	1513	2039.93	0.07
H-STATUS:WORDS	1	10.01	1501	2023.84	0.02
D-STATUS:WORDS	1	6.76	1500	2017.08	0.01

Table 6. Linear model including all downstepped contours (All-DS)

Predictor	**DF**	**Deviance**	**Residual DF**	**Residual Dev.**	**Pr(CHI)**
NULL			1550	2149.60	
SBEG	1	6.32	1549	2143.28	0.01
SF	1	0.49	1548	2142.79	0.48
H-STATUS	3	11.69	1545	2131.10	0.01
D-STATUS	2	6.48	1543	2124.62	0.04
SPEAKER	3	22.06	1540	2102.56	0.00
ENTITIES	1	19.18	1539	2083.38	0.00
WORDS	1	1.32	1538	2082.06	0.25
SF:H-STATUS	3	7.92	1532	2072.77	0.05
SF:SPEAKER	3	13.15	1524	2057.45	0.00
D-STATUS:ENTITIES	1	2.74	1509	2042.83	0.10
D-STATUS:WORDS	1	3.54	1500	2034.71	0.06

When we examine All-DS, a similar analysis shows that topic beginning position, Hearer status, Discourse status, speaker identity, and number of discourse entities within phrase are all significant predictors of downstep in the phrase. There are significant interactions between segment finality and Hearer status and between segment finality and speaker identity, with tendencies to significance between Discourse status and length of NP in words and between Discourse status and length of NP in discourse entities. Since these findings show similar, but more definite tendencies than those in Table 5, we may tentatively conclude that the relationships between

downstepped contours, discourse structure, and the given/new distinction posited in the literature for the DS contour (H* !H* L- L%) may be applied more broadly to downstepped contours in general.

4.2. Discussion

This corpus-based study of downstepped contours has examined two discourse functions of downstepped contours hypothesized in the literature, a topic structure marking function and a given information marking function. We have found evidence that the particular category of DS contours (marked by !H* pitch accents only) and, more broadly, the set of all downstepped contours (All-DS), do appear to serve at least two functions.

These contours are indeed associated with key aspects of discourse topic structure, particularly serving to mark topic ending phrases. However, they are clearly used in different proportions and even for different functions by different speakers: For example, only one of our speakers appears to employ them routinely in topic beginning position in read speech and there is considerable variation among the other speakers. Downstepped contours are frequently used over both Given and New NPs; there appears to be no simple association between downstep and givenness, although phrases that are Hearer-Inferable are more frequently uttered with downstep than with other contours. This observation supports Pierrehumbert and Hirschberg 1992's proposal that downstepped pitch accents convey that information should be inferable from Speaker and Hearer's shared beliefs. DS contours **are** used much more frequently that fully deaccented contours when material in an NP represents all-Given information. When we examine all downstepped contours together, we find that 50% of Hearer-Given NPs, 57% of Hearer-Inferable NPs, and 50% of Discourse-Given NPs where all discourse entities are given, are uttered with a downstepped contour. However, 56% of Hearer- and Discourse-New NPs are also uttered with downstep. The reason why some given NPs are deaccented and others uttered with downstep still remains elusive, despite the evidence of the more frequent occurrence of the latter. While this phenomenon may well represent a constraint on accenting too many items in an NP (Ladd 1996), the choice of when to use downstep and when to use another contour remains to be determined.

Our more general modeling of downstep in the context of its potential predictors, including discourse structure, information status, speaker variability, and simple features such as NP length, has been shown to support relationships among downstep and each of these variables. Further research will be needed to test which other factors play a role in the choice of these contours.

Acknowledgements

This research was partially supported by a grant from the National Science Foundation IIS-0307905, "Dialogue Prosody in Interactive Voice Response Systems". Thanks to Barbara Grosz and Christine Nakatani for their work in collecting and annotating the BDC in collaboration with the first author of this paper. Thanks also to Jennifer Venditti Ramprashad for fruitful discussions on the uses of DS contours, particularly with respect to the importance of information status.

Notes

1. Here and throughout this paper we will identify intonational phenomena using the ToBI labeling scheme for SAE (Silverman et al. 1992, Pitrelli et al. 1994).
2. The Boston Directions Corpus was designed and collected in collaboration with Barbara Grosz and Christine Nakatani at Harvard University.
3. A fuller description of the ToBI system may be found in the ToBI conventions document and the training materials available at http://ling.ohiostate.edu/tobi.
4. The annotation guide presents the idea that natural fluent speech comprises phrases organized into coherent units or chunks and introduces the term 'discourse segment' to refer to these chunks. It then introduces the notion of the reason or purpose that underlies a speaker's saying something, and the term DSP to refer to these purposes. The task of segmenting a discourse is described by analogy with outlining, but special attention is paid to specific differences between these two kinds of labeling. Examples of recipe descriptions are used to explain segments and different kinds of relationship between them are given. Subjects are instructed that the primary question they should ask when segmenting is "Why did the speaker say <phrase>?" The "why's" form the basis of descriptions of the DSPs. Relationships among the various "why's" determine how phrases are chunked into segments.
5. For earlier studies, interlabeler reliability was calculated using Cohen's (1960) coefficient. This measure factors out chance agreement, taking the expected distribution of categories into account. For earlier studies, expected agreement was calculated based on the distribution of e.g. SBEG versus non-SBEG labels

for all labelers on one of the nine direction-giving tasks. Using these distributions, kappa coefficients for each pair of labelers were calculated for the remaining eight tasks in the corpus and kappas averaged over the pairs. Typically, kappa values of.7 or higher provide evidence of good reliability (Carletta 1996). Expert labelers achieved average kappa scores of .8 on the marking of SBEG in the spontaneous speech of Speaker 1, the only speaker they annotated.

References

Beckman, Mary E., Julia Hirschberg, and Stephanie Shattuck-Hufnagel
2004 The original ToBI system and the evolution of the ToBI framework. In *Prosodic Models and Transcription: Towards Prosodic Typology*. Sun-Ah Jun (ed.), Chapter 2, 9-54. Oxford: Oxford University Press.

Carletta, Jean
1996 Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics* 22(2): 249-254.

Cohen, J.
1960 A coefficient of agreement for nominal scales. *Educational And Psychological Measurements* 20: 27-46.

Dahan, D., M. K. Tanenhaus, and C. G. Chambers
2002 Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47: 292-314.

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

Hastie, Helen Wright, Rashmi Prasad, and Marilyn Walker
2002 What's the trouble: Automatically identifying problematic dialogues in DARPA Communicator dialogue systems. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Philadelphia, 384-391.

Hirschberg, Julia, and Christine Nakatani
1996 A prosodic analysis of discourse segments in direction-giving monologues. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 286-293, Santa Cruz.

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

Hirschberg, Julia, and Owen Rambow
2001 Learning prosodic features using a tree representation. *Proceedings of EUROSPEECH 2001*, 1175-1180, Aalborg.
Hirschberg, Julia, 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.
Ladd, D. Robert
1979 Light and shadow: A study of the syntax and semantics of sentence accents in English. In *Contributions to Grammatical Studies: Semantics and Syntax*. L. Waugh Baltimore and Frans van Coetsem (eds.), 91-131, University Park Press.
1996 *Intonational phonology*. Cambridge: Cambridge University Press.
Liberman, Mark and Janet Pierrehumbert
1984 Intonational invariants under changes in pitch range and length. In *Language Sound Structure*, Mark Aronoff and Richard Oehrle (eds.), Cambridge: MIT Press.
Nakatani, Christine, Barbara Grosz, and Hirschberg Julia
1995 Discourse structure in spoken language: Studies on speech corpora. *Proceedings of the American Association for Artificial Intelligence Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, Stanford, March.
Nickerson, J. S., and J. Chu-Carroll
1999 Acoustic-prosodic disambiguation of direct and indirect speech acts. *Proceedings of the XIV International Congress of Phonetic Sciences*, 1309-1312, San Francisco.
Pierrehumbert, Janet B.
1980 The Phonology and Phonetics of English Intonation. Ph.D. diss., Department of Linguistics, Massachusetts Institute of Technology.
Pierrehumbert, Janet, and Julia Hirschberg
1990 The meaning of intonational contours in the interpretation of discourse. In *Intentions in Communication*, Phil Cohen, Jerry Morgan, and Martha Pollack, (eds.), 271-311, Cambridge: MIT Press.
Pitrelli, John, Mary Beckman, and Julia Hirschberg
1994 Evaluation of prosodic transcription labeling reliability in the ToBI framework. *Proceedings of the International Conference on Spoken Language Processing*, 2: 123 -126, Yokohama.
Price, Patti J., Mari Ostendorf, Stephanie Shattuck-Hufnagel, and C. Fong

1990 The Use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90.

Prince, Ellen F.

1981 Toward a taxonomy of given-new information. In *Radical Pragmatics*, Peter Cole, (ed.), 223-255, New York: The Academic Press.

1992 The ZPG letter: Subjects, definiteness, and information-status. In *Discourse Description: Diverse Analyses of a Fund Raising Text*, Sandra Thompson and William Mann, (eds.), 295-325, Philadelphia: John Benjamins B. V.

Sag, Ivan A. and Mark Y. Liberman

1975 The Intonational disambiguation of indirect speech acts. *Papers from the Eleventh Regional Meeting of the Chicago Linguistic Society*, 487-497.

Silverman, Kim, Mary Beckman, Janet Pierrehumbert, Mari Ostendorf, Patti Price, and Julia Hirschberg

1992 ToBI: A Standard Scheme for Labeling Prosody. Proceedings of ICSLP 1992, Banff, 867-879.

Talkin, David

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

Venditti-Ramprashad, Jennifer

2002 Personal Communication.