Intonation and discourse processing

Jennifer J. Venditti and **Julia Hirschberg**
Columbia University
{jjv,julia}@cs.columbia.edu

Abstract

This paper describes intonational cues to discourse structure, and the role that intonation plays in spoken discourse processing. We begin by discussing two main structures in discourse that one must consider when doing research on discourse processing: *segmentation* and *information status*. We then review a number of key studies from the phonetics literature which have investigated the intonational marking of these structures. Next, we discuss in detail the psycholinguistic research to date which has examined the role that intonation can play in facilitating or inhibiting the processing of discourse in English and other related languages. We conclude by outlining directions for future research in the area of intonation-discourse processing.

1 Introduction

Intonation is an integral part of every spoken language utterance. It can provide cues to the linguistic structure of the speaker's message, her emotional state, or her communicative intent. Despite this wealth of information available in the signal, surprisingly little is known about how listeners might go about integrating this information into their interpretation of an utterance. The goal of this special session is to discuss the contributions of intonation to spoken language processing. In their paper, Speer et al. [39] describe how intonation can be used in the parsing of syntactic structure. Our paper will concern the processing of another type of linguistic structure, namely the discourse structure.[1]

In order to talk about the role of intonation in discourse processing, we must first clarify what we intend by the terms *discourse* and *processing*. By *discourse*, we mean not only aspects of linguistic structure above the sentence level, such as paragraph or topic structure and dialog turns, but also the dynamic shifts in information status, including salience, focus of attention, and the *given/new* distinction. In Section 2 we will outline a model of discourse structure, proposed by Grosz & Sidner [18], which incorporates these two aspects into a unified account of discourse structuring. The term *processing* can mean a variety of things, from off-line comprehension of and judgments about an utterance's structure and meaning, to the on-line moment-by-moment interpretation of that utterance, to the implicit 'workload' that is associated with such interpretation. We will attempt to address all of these aspects of processing in our review of the literature. In Sections 3 and 4 we will discuss key studies which have investigated the intonation-discourse interface. We will include selected works from the vast phonetics literature on this topic, and integrate it with the small but growing literature from the psycholinguistics community on the role of intonation in discourse processing. We will conclude in Section 5 by offering a few suggestions for areas of research that we find particularly intriguing and fruitful for future research on intonation and discourse processing.

[1] For issues concerning the parsing of intonational structure itself, see Beckman [3]. Readers are also referred to Cutler et al.'s [12] extensive review of prior literature on intonation in spoken language understanding, including discourse understanding.

2 Discourse Structuring

Research on the intonation-syntax interface has shown that intonation can play an important role in cueing structures such as clause boundary location, PP and relative clause attachment, and NP bracketing (e.g. *old men and women*). In the discourse domain, what structures are important in processing, where do the potential ambiguities lie, and what intonational cues may potentially disambiguate them? In this section we will discuss two main aspects of discourse structuring: (i) *segmentation* and *hierarchy*, and (ii) *information status*, including salience, focus of attention, and *given/new* information. While there are many theoretical accounts of these phenomena, we will focus here on how these are represented in the theory of discourse structuring and coherence proposed by Grosz & Sidner [18], which has been widely used in both computational and experimental research (e.g. [29, 16, 24]).

2.1 Segmentation

Most researchers now assume that a spoken (or written) discourse is more than just a string of utterances, but that individual utterances of a discourse are grouped into higher-level units. In order to characterize these units, previous studies have used written cues such as 'paragraphs', or labelers' or researchers' intuitive notions of topic structure in the data under analysis. Often, the discourse structure description is study-specific and cannot be generalized to other data. What

is generally lacking is an independently-motivated theory of discourse structuring, which can be empirically determined (by trained labelers) in a reliable manner.

One such independent theory is the intention-based proposal by Grosz & Sidner [18]. Under this proposal, utterances are grouped into cohesive units known as *discourse segments* (DS), which serve as the building blocks that make up the discourse. Utterances grouped in a DS share a common property: they all contribute to the overall *purpose* or *intention* that a speaker has for producing that particular segment. The purposes of the segments (*discourse segment purposes* or 'DSPs') then contribute to the overall purpose of the discourse (the *discourse purpose* or 'DP'). In other words, a speaker generally has a reason for producing a discourse. Individual utterances contribute to the DSPs of the segments to which they belong, which in turn contribute to the overall DP. In addition, in this theory a DS is related to other DSs in one of two ways: by *dominance*, a hierarchical relationship in which the purpose of the dominated segment contributes to the purpose of the dominating segment) or by *satisfaction-precedence*, a linear relationship in which one DSP must be satisfied **before** the DSP of another [18, 30].

This theory of discourse organization has been put to the test by asking human labelers to segment speech corpora using the segmentation guidelines developed by [30]. Such studies have demonstrated a high degree of inter-labeler reliability [17, 21, 29]. Research using this method of segmentation has shown that DS boundaries may be marked by linguistic means such as specific lexical items known as *cue phrases* (e.g. *so*, *next*, *finally*, etc.) [22], or shifts in tense, but that such cues may not always be present. In Section 3 we will summarize recent research investigating intonational cues to such intention-based segmentation.

2.2 Information Status

As with discourse segmentation, numerous theoretical constructs have been proposed in the literature to capture the notion that discourse entities change their status over the course of a discourse, from new information to old (or *given*) information, from focus to background. These changes in information status are closely related to the *accessibility* of individual entities as they are referred to by discourse participants. The notions of *given* vs. *new* (see e.g. [19, 8, 9, 10, 34]), whereby entities newly introduced into the discourse are considered *new*, and those already in the discourse context are *given*, while widely invoked in both theoretical and psycholinguistic research, are notoriously difficult to define. What does it mean for a *given* entity to be 'already in the discourse context'? A number of solutions to this problem have been proposed, including Halliday's notion of 'recoverability' [19], Chafe's definition with respect to the listener's 'consciousness' [8, 9], and Prince's multi-dimensional taxonomy [34, 35]. Here, we choose to focus on how *given/new* might be represented in Grosz & Sidner's theory of intention-based discourse structuring.

As mentioned in the previous section, a discourse is composed of a number of discourse segments (DS), each with its own purpose. This linguistic structure interfaces with another important structure, namely the *attentional state* of the discourse. According to Grosz & Sidner's proposal, the onset of each DS opens up a new *global focus space* in the ever-evolving record of speaker and hearer's attentional state, to which discourse entities may be added as they are referred to. For example, if a speaker utters *Now I will build a house*[2], the cue word *now* signals that a new DS has begun, and a corresponding focus space is added to the representation of the attentional state, into which the discourse entity representing *house* will be inserted. Under this approach, when an entity such as *house* is first added to discourse's focus space, we might say that it is also considered *new* information. Once introduced, a now *given* entity may remain *salient* and *accessible* in the discourse, or it may lose its salience and accessibility as the discourse proceeds [18, 29, 45]. Thus, we may view Grosz & Sidner's notion of *attentional focus* as modeling two kinds of information status: both the *given/new* distinction and the notion of *salience* or *accessibility*. So in this way, *given* information is no longer defined as 'mentioned within the last N utterances', but rather is directly related to an entity's status with respect to the dynamic representation of attentional state. In Section 4 we will describe how this approach to modeling information status can account for the distribution of intonational prominences in spoken discourses.

3 Intonation in Discourse Segmentation

There is considerable evidence in the literature that intonational features can signal the structuring of utterances into larger discourse segments.[3] In an early study, Lehiste used written 'paragraphs' as the discourse unit of interest [26]. She found that English utterances with high F0 peaks are perceived by listeners as being paragraph initial. Paragraph-medial and final utterances tend to have lower F0 peaks. In another study, Silverman found that manipulating the pitch range of intonation phrases in English using resynthesis can cause listeners to segment discourses with ambiguous structures differently: phrases with an expanded pitch range are likely to be judged as paragraph initial,

[2] See discussion of house-building experiments by Terken [42] and Swerts & Geluykens [40] in Sections 3 and 4.1, respectively.

[3] In this paper we will focus on the use of fundamental frequency (F0) in cueing discourse structure. Readers are referred to the literature for descriptions of other acoustic cues that may be used. Also, our focus will primarily be on English and languages with similar intonational systems (e.g. Dutch). Studies of the intonation-discourse interface in other languages (e.g. Japanese, see [45, 47]) have had similar findings, though the intonational means used to cue discourse structures are different.

while final lowering can cue paragraph finality [37].

Other studies have defined discourse units in terms of 'topic' structure: stretches of speech in which the speaker is mainly discussing a single entity. Yule suggested that intonation can be used to mark the boundaries of topic units in English spontaneous speech: a structure which he termed the 'paratone' [49]. Swerts & Geluykens examined topic units in Dutch, defined in their house-building task as a stretch of speech in which a specific piece of the house is being described [40]. They also found that F0 is high at the beginning of such units, and gradually declines to the unit end.

In task-oriented discourses and similar genres, the notion of 'topic units' may be equated to the intention-based discourse segments defined by Grosz & Sidner [18, 30]. For example, in Swerts & Geluykens's [21] house-building task, the topic unit which describes the construction of the *front door* is likely to be the same as the intention-based DSP 'Tell the listener how to construct the front door' in the Grosz & Sidner framework. An advantage of Grosz & Sidner's intention-based discourse segmentation scheme is that it can describe other discourse genres as well. The relation of intonation to discourse segmentation within this framework has been studied extensively by Hirschberg and colleagues (e.g. [23, 17, 21, 29]), in various discourse genres. Hirschberg found that increased F0 values (both maximum and mean F0) are characteristic of (intermediate) prosodic phrases which labelers agree to be discourse segment initial, relative to other phrases in the database. Likewise, lower F0 values relate to DS-medial and DS-final judgments [17, 21].[4]

Given that a number of phonetic studies, using a variety of theoretical constructs, have found that discourse segmentation can be cued by intonational means, the question of most relevance to psycholinguists is to what extent this segmentation can occur on-line, as the listener processes the incoming spoken discourse. Finding that, for example, DS-initial phrases are uttered with a higher F0 in comparison with other phrases can be a tremendous aid in speech synthesis, where such phrases can be systematically distinguished from an otherwise 'default' F0 topline (see e.g. [23]). However, such a finding does not necessarily shed light on whether listeners (or real-time automatic recognition systems) can use this same intonational information in on-line discourse segmentation. In order to address this concern, Hirschberg & Nakatani examined the *relative change* in F0 (and other parameters) over a local window of two consecutive phrases [21]. They found that the previously-reported effects of DS position are true even at this more local level. That is, there is a significant increase in F0 change from one phrase to the next when the second phrase is DS-initial, in comparison to when that phrase is medial or final. In addition, DS-medial phrases are marked by an increased F0 change in comparison with final phrases. These findings suggest that discourse segmentation could in many cases be accomplished on-line as the discourse unfolds, by examining the local change in overall phrasal pitch range from one intonation phrase to the next.

[4]These studies documented a number of other acoustic-prosodic features which are also reliably related to DS position, such as amplitude, speaking rate, and pause durations.

4 Intonation and Information Status

There has also been considerable research on the role intonation can play in cueing information status. Speakers can indicate the salience or accessibility of a discourse referent by varying the intonational prominence of referring expressions. This is accomplished by *pitch accents* in languages such as English and Dutch, though other languages may use different means. (For example, see research by Kang [25] on Korean and Venditti and colleagues [45, 47] on Japanese suggesting that local *pitch range* and/or *phrasing* variations can cue information status in these languages.)

4.1 Intonational Marking

The *given/new* distinction often cited in the discourse literature was defined by Halliday directly in terms of the speaker's choice of intonational form [19]. For Halliday, *new* information is focal information which "the speaker presents ... as not being recoverable from the preceding discourse" (regardless of whether or not it had been mentioned before) [19, p. 204], and is marked in English by a 'tonic' or 'nuclear' pitch accent. While Halliday's claim is that the *given/new* distinction is defined solely by the speaker's choice of intonational grouping and prominence, subsequent studies have attempted to relate the intonational phenomena to independent text-based characterizations of *given* vs. *new* information. For example, Brown [6] used Prince's [34] taxonomy of discourse givenness to describe variations in intonational prominence in English task-oriented speech. She found that speakers tend to place pitch accents on *new* information, while marking *given* information by deaccenting. However, Brown also points out an instance in her data in which a *given* entity is re-introduced into the discourse after some digressions, and is marked by a pitch accent.[5] Using only Prince's taxonomy of givenness, along with a direct mapping of these categories to intonational prominence markings, Brown cannot account for such accenting of re-introduced entities. However, approaches using the notion of a cache/buffer of a fixed number of utterances may be able to capture such phenomena. That is, the entity is no longer salient if the number of utterances defined by the cache size have intervened. But what cache size is appropriate? And is the same size appropriate for all discourse situations?[6]

[5]The accenting of re-introduced entities has also been observed by Hirschberg [20].

[6]Cf. Cahn's [7] recent work on memory-based salience.

The use of topic-based discourse segmentation is one way to better define what it means for an entity to be *given*. Terken [42] examined accent distribution in Dutch house-building monologues, using a topic unit defined as a stretch of speech in which a specific piece of the house is being described (see also Swerts & Geluykens [40] mentioned above). He found that both topics and non-topics are newly introduced using accented full NPs (97% and 81%, respectively). This is consistent with the accenting of *new* entities. However, Terken observed that the realization of later mentions (within the topic unit) depends on the topic status of the entity: topics are mainly realized by unaccented pronouns (51%), but accented and unaccented full NPs are also found (33% and 5%, respectively). Later mentions of non-topics, on the other hand, are primarily realized by accented full forms (74%), though unaccented full forms exist as well (18%). These results suggest that while there is a general relationship between *given/new* (as defined by topic unit segmentation) and pitch accenting, there are additional factors which also affect accent distribution. We will return to this issue below.

Hirschberg & Pierrehumbert [23] suggest that the notions of *given* and *new*, and their relation to pitch accenting, can be explained by a model of global attentional salience such as that proposed by Grosz & Sidner [18]. Working within this framework, Nakatani [29, 28] observed that entities which are first introduced into the current global focus space (which models the current intention-based DS) tend to be realized with accented referring expressions, while those entities already existing in the space (and hence globally salient) tend to be realized with unaccented expressions. Nakatani also notes that "references to entities that are either in a neighboring focus space on the focus stack, or in the most recently popped focus space, [also] do not require accentual prominence" [28, p. 149].

To the extent that intention-based discourse segmentation may in many cases correspond closely to the topic-based segmentation in Terken's house-building discourses, Terken's results can be directly interpreted in terms of this new approach. In addition, Nakatani's observations using Grosz & Sidner's model can account for two of Terken's 'exceptions': reference to the entity *house* using a non-prominent expression, and decaccenting of some referents when the antecedent is in the previous topic unit. In the first case, Terken notes that "expressions referring to the house itself are often deaccented, even though the house has not been mentioned over long stretches of discourse" [42, p. 280]. One possible explanation for this is that the entity *house* could reside in the global focus space in the representation of the discourse's attentional state, for example, due to its mention in a superordinate DS whose purpose is to *Explain how to assemble the house*. If this is the case in Terken's data, then Grosz & Sidner's model of global focus would characterize this entity as being in non-immediate global focus, and this would license the use of an unaccented expression to refer to the *house* in subsequent embedded segments. In the second case of 'exceptions', Terken observes deaccenting of some referents across topic unit boundaries. Although Terken does not describe these exceptions in full detail, it seems that they could be an instance of the same phenomenon observed by Nakatani [29] (and independently by Grosz & Sidner [18] and Davis & Hirschberg [15]) — namely, that an entity in a just-completed (or 'popped') DS can be still salient and thus does not need to be accented when mentioned in the next sister DS. This model of salience and accessibility could also explain the apparent 'exception' noted independently by Brown [6] and by Hirschberg [20], that *given* entities which are re-introduced into the discourse are marked by pitch accents. If the previous mention of the entity occurred in a non-adjacent and non-embedding DS, this would warrant re-introduction using a pitch accent under this account. Thus, Grosz & Sidner's dynamic model of global focus driven by intention-based discourse structuring can provide a rich architecture in which to examine patterns of accentuation in both naturally-occurring and experimental data.

4.2 Processing

While information status has been shown in a number of studies to strongly influence pitch accent distribution in languages such as English and Dutch, to understand the role of intonation in spoken discourse processing, we must also investigate whether listeners are in fact sensitive to such markings. In this section we discuss studies which suggest that accentuation does indeed play a role in processing.

An early study by Most & Saltz [27] asked listeners to choose which of two *wh*-questions an intoned target answer would be an appropriate reply to. They found that listeners' choice of matching questions was related to the accentuation in the target answer. For example, an answer such as *The MECHANIC fixed the car* was taken to be the answer to *Who fixed the car?*, rather than to *What did the mechanic fix?*. Birch & Clifton [4] also examined the effect of accentuation in processing question-answer pairs. They used both 'makes sense' judgments (i.e. listeners provided speeded judgments of whether an answer made sense given the question) and prosodic appropriateness ratings of answers with varied accentuation patterns. They found that answers in which *new* information was accented and *given* information deaccented were not only rated as more appropriate by listeners, but were understood more quickly in speeded judgments.

In another study, Bock & Mazzella [5] used a comprehension time paradigm to determine the effect of appropriate accentuation on processing of so-called denial-counterassertion pairs such as *Arnold didn't fix the radio. Doris fixed the radio.* They found that comprehension times of the target counterassertions were

shorter when focal (i.e. *new*) information was accented, compared to non-focal information, suggesting that appropriate accentuation facilitates comprehension in these utterances. More recently, Davidson [14] used a phoneme-monitoring paradigm to demonstrate that listeners use accentuation patterns in denials to direct attention to alternatives presented in the counterassertion, consistent with Bock & Mazzella's findings.

Terken & Nooteboom [44] have also demonstrated that listeners expect *new* information to be pitch accented and *given* information to be deaccented (see also [31]). Inappropriate accentuation on target words slowed verification latencies in these experiments: that is, accenting *given* information slowed reaction times, as did deaccenting *new* information. In their experiments, Terken & Nooteboom defined *given* information as an NP whose antecedent was in the same grammatical role either (a) in the immediate preceding utterance only, or (b) in a number of preceding utterances. The effect of accentuation on the processing of *given* information was the same for either definition of *given*. Under the model of attentional focus described in Sections 2.2 and 4.1 above, both definitions would predict that the *given* NP is in global focus, consistent with the findings of production studies using this framework.

Many of the processing studies examining the relationship between accentuation and information status have used experimental paradigms that probe so-called 'off-line' comprehension. A very recent study by Dahan et al. [13] suggests that accentuation also affects referential interpretation at very early stages of processing, even before the entire target word has been heard. Dahan et al. used eye-tracking to monitor listeners' fixations on pictured entities as they heard simple pre-recorded instructions to manipulate the entities on a computer screen, as in *Put the candle/candy below the triangle. Now put the CANDLE/candle above the square*. Eye-tracking has become a popular methodology for investigating real-time spoken language processing, since it allows experimenters to monitor (visual) attention to referents without interrupting the speech stream (unlike the gating paradigm, for example), and because eye movements have been found to be closely time-locked to the auditory input in such tasks (see [41] for a brief introduction to the eye-tracking paradigm). In Dahan et al.'s study, the visual scene contained various objects, two of which shared the same primary-stressed first syllable (e.g. *candle* and *candy*). The first part of the auditory instruction introduced either the *candle* or the *candy* into the discourse context (see example above), establishing it as *given*. The second part of the instruction then referred to the *CANDLE/candle* using either an intonationally prominent or non-prominent surface form.[7] Note that the target noun in the second instruction is (crucially) temporarily ambiguous during the first syllable [kæn], and thus both *candle* and *candy* are potential referents at this stage. Dahan et al.'s results showed clear effects of accentuation on reference resolution. They found that, while listeners eventually did fixate on the target noun (i.e. *candle*), which was uniquely identifiable in its full form, fixations on the competitor (i.e. *candy*) differed significantly depending on accentuation: when *candle* had been mentioned in the first instruction, there were more fixations on *candy* when [kæn] was accented. Likewise, when *candy* was previously mentioned, there were more fixations on *candy* when [kæn] was deaccented. That is, listeners took accented [kæn] to refer to *new* information, and deaccented [kæn] to refer to *given* information, even at the earliest stages of lexical access. This confirms that accentuation can indeed be reliably used by listeners to process discourse representations, both in global (off-line) comprehension, as well as on-line, as a discourse is unfolding.

4.3 Property-sharing Constraints

A discussion of the relation between information status and accentuation would not be complete without mention of one factor which has come to the forefront in recent studies: the fact that *given/new* interacts with *property-sharing constraints* in the distribution of accents in discourse. In many previous studies, this factor has either been overlooked or implicitly controlled for (e.g. by placing target and antecedent in the same grammatical role). In a few studies, this factor has been systematically varied, with revealing results.

Terken & Hirschberg [43] examined the distribution of accents in elicited spontaneous descriptions, and found that prior mention (even in the immediately preceding utterance within the same discourse segment) is not a sufficient predictor of deaccenting. The target and antecedent must in addition share the same grammatical role to warrant deaccenting (see also [32]). This importance of property-sharing (here, grammatical role) was also demonstrated by one of the experiments reported by Dahan et al. [13]. Using the same eye-tracking paradigm and experimental task as their study reported above, they examined instruction sequences in which the target and antecedent did not share the same grammatical/thematic role, as in *Put the necklace below the candle. Now put the CANDLE above the square*. Analysis of eye fixations revealed that there were no competitor effects in this condition, as were observed in the conditions in which the target and antecedent shared the same grammatical role. In other words, the accented [kæn] was immediately interpreted as referring NOT to *new* information (which would have led to fixations on the competitor *candy*), but to *given* information which was realized in a **different** (non-focused) grammatical position.

[7] In Dahan et al.'s stimuli, prominent expressions were marked by H* or L+H* pitch accents, and non-prominent (or 'deaccented') expressions were marked by downstepped H+!H* (and not by total deaccenting). Pierrehumbert & Hirschberg [33] and Ayers [2] have also observed that downstepped accents may sometimes be functionally similar to deaccenting.

These and other studies show that the distribution of accents in discourse depends on more than just the *given/new* distinction. Not only are the notions of *given/new* notoriously tricky to define, but even in the clearest cases of *given* referring to information mentioned in the immediately previous utterance and *new* referring to unmentioned information, there are other constraints such as sharing of grammatical/thematic role or surface position that complicate matters. Another factor which Terken & Hirschberg briefly touch upon but reserve for future research is the possibility that this observed persistence of grammatical role "may arise only due to the syntactic parallelism of successive utterances in [their] context and target utterances" [43, p. 142] — perhaps a more restrictive notion of 'property-sharing'.

In a series of recent eye-tracking studies, Venditti et al. [46, 48] demonstrated that syntactic parallelism has a significant effect on the interpretation of (ambiguous) nuclear-accented pronouns. They found that while accented pronouns serve to shift reference in parallel constructions, such as in the now infamous *John hit Bill and then HE hit George*, listeners had difficulty interpreting accented pronouns in non-parallel sequences (e.g. *John hit Bill and then HE ran away*). Moreover, even in parallel constructions, significant preference for switched reference (as indicated eye fixations) only emerged after listeners had heard the (identical) verb, which provided strong evidence for syntactic parallelism. Since the auditory stimuli in Dahan et al.'s [13] study (see examples above), and the spontaneous productions in Terken & Hirschberg's [43] study (e.g. *The ball touches the diamond. The ball touches the star.*) involved not only sharing of grammatical role but also syntactic parallelism, more research is needed to clarify which of these factors (or both) are responsible for the observed patterns/effects of accentuation, as Terken & Hirschberg point out.

5 Future Directions

In this paper, we have summarized the current state of knowledge of the intonation-discourse interface, and the role intonation can play in discourse processing. There is much more work to be done. In this section we outline a number of intriguing directions open for future research in this area.

Processing in dynamic models of attention and discourse salience. Much of the discourse processing literature has focused on the *given/new* distinction as defined by adjacent utterance pairs. We have briefly described Grosz & Sidner's [18] model of intentional structuring and attentional focus which has been used in a number of production studies examining the intonation-discourse interface. Such a model will allow future studies to investigate the more dynamic aspects of processing across extended stretches of discourse.

Property-sharing, parallelism, and other constraints on accentuation. Although a number of studies have observed that *given* information often does bear intonational prominence, the extent to which factors such as property-sharing or parallelism (however defined) can explain accent distribution has yet to be fully investigated in either production or processing studies. Other factors that must also be addressed with respect to this issue include: the asymmetry between nuclear and pre-nuclear accents in marking *given* information, and the functional similarity of downstepped accents and deaccenting (see e.g. [13, 33, 2]).

The time course of integration of intonational information. Most of the intonation-discourse processing studies to date have involved off-line comprehension or appropriateness judgments. Notable exceptions are Dahan et al.'s [13] and Venditti et al.'s [46, 48] experiments using eye-tracking, which were able to probe on-line integration of intonational information as a discourse unfolds in real-time. More studies are needed to investigate the exact time course of such information integration. A number of previous studies have suggested that intonational information occurring even prior to the event of interest can be used by listeners in processing. For example, Cutler [11] showed that an intonation contour leading up to a target word which was consistent with that word being accented (although the word itself was a neutral version spliced in) resulted in a phoneme-monitoring response time advantage. In Terken & Nooteboom's study [44] described above, they suggest that an effect of grammatical role on observed reaction times might be due to facilitation by the preceding intonation contour. Since their predicate target NPs were all preceded by a falling pitch movement (which in Dutch typically signals that the remaining portion of the utterance contains non-focal information), they suggest that listeners were likely able to identify the predicate NP as deaccented (thus *given*) **before** the NP itself was even uttered. These and other studies (see e.g. Bock & Mazzella's [5] and Davidson's [14] discussions of accentuation facilitating comprehension of subsequent information in denials) underscore the importance of more research on the time course of integration of anticipatory and other intonational cues in real-time discourse processing.

Cross-linguistic perspectives on intonation and discourse processing. Our discussion here has focused primarily on the role of intonation in processing English and Dutch discourses. Research shows that in these languages, variations in pitch range can aid in discourse segmentation, and pitch accent distribution has a strong influence on the processing of information status. Much work needs to be done to investigate how these and other discourse structures are cued in languages which don't have pitch accent systems like those of English and Dutch. For example, Kang [25] has observed that speakers can use pitch range and

accentual phrasing (among other acoustic features) to mark information status in Korean discourses. Venditti and colleagues [47, 45] also found that systematic variation in pitch range can mark intention-based structuring, information status, and topic transitions in Japanese. The next step is to extend these production results to studies of spoken discourse processing in a range of languages with varied intonation systems.

The role of ambiguity 'awareness' and experimental design. The many experimental studies reviewed here have demonstrated that intonation can play a major role in discourse processing, both in discourse segmentation as well as processing information status. However, most of these studies have involved highly-structured laboratory experiments, or have examined the speech of trained speakers — two factors which may have inadvertently inflated the influence of intonation. In their paper in this special session, Speer et al. [39] describe recent studies on syntactic processing which have questioned the extent to which intonational cues are reliably produced or used by naïve speakers/listeners. For example, Allbritton et al. found that speakers could only reliably produce prosodic cues to disambiguate certain syntactic structures when the ambiguity was pointed out [1]. Snedeker & Trueswell also found that speakers only provided reliable intonational cues when they were aware of the ambiguity, but did not produce disambiguating cues when the two competing structures were manipulated in a between-subjects design (that is, a given subject only encountered one version of the structure) [38]. In contrast, Speer et al. report that naïve subjects could indeed produce reliable cues in their experiment, even when the structure was unambiguous [39]. In the intonation-discourse domain, the effects of speaker/listener 'awareness' or experimental design (e.g. using between- vs. within-subjects designs, including sufficient distractor trials, etc.) on the production/perception of intonational cues have yet to be formally examined.[8] Clearly then, more research is needed to determine the role of intonation in processing of naturally-occurring discourse, by naïve speakers and listeners, using experimental designs which do not highlight potential ambiguities.

References

[1] D. Allbritton, G. McKoon and R. Ratcliff, "Reliability of prosodic cues for resolving syntactic ambiguity," *Jorurn. of Exper. Psych.: Learning, Memory and Cognition* 22: 714–735, 1996.

[2] G. Ayers Elam, *Nuclear accent types and prominence: Some psycholinguistic Experiments*, Ph.D. thesis, Ohio State University, 1996.

[3] M.E. Beckman, "The parsing of prosody," *Language and Cognitive Processes* 11: 17–67, 1996.

[4] S. Birch and C.E. Clifton, "Focus, accent, and argument structure: Effects on language comprehension," *Language & Speech* 38: 365–391, 1995.

[5] J.K. Bock and J.R. Mazzella, "Intonational marking of given and new information: Some consequences for comprehension," *Memory and Cognition* 11: 64–76, 1983.

[6] G. Brown, "Prosodic structure and the given/new distinction," In D.R. Ladd and A. Cutler (eds.), *Prosody: Models and Measurements*, Springer-Verlag, pp. 67–78, 1983.

[7] J.E. Cahn, *A Computational Memory and Processing Model for Prosody*, Ph.D. thesis, MIT, 1998.

[8] W.L. Chafe, "Language and consciousness," *Language* 50: 111–133, 1974.

[9] W.L. Chafe, "Givenness, contrastiveness, definiteness, subjects, and topics," In C.N. Li (ed.), *Subject and Topic*, Academic Press, pp. 27–55, 1976.

[10] H.H. Clark and S.E. Haviland, "Comprehension and the given-new contract," In R.O. Freedle (ed.), *Discourse Processes: Advances in Research and Theory (vol. 1)*, Ablex Publ., pp. 1–40, 1977.

[11] A. Cutler, "Phoneme-monitoring reaction time as a function of preceding intonation contour," *Perception and Psychophysics* 20: 55–60, 1976.

[12] A. Cutler, D. Dahan, and W. van Donselaar, "Prosody in the comprehension of spoken language: A lit. review," *Lang. & Speech* 40(2): 141–201, 1997.

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

[14] D.J. Davidson, *Association with focus in denials*, Ph.D. thesis, Michigan State University, 2001.

[15] J.R. Davis and J. Hirschberg, "Assigning intonational features in synthesized spoken directions," *Assoc. for Comp. Ling.*, 1988, pp. 187–193.

[16] P.C. Gordon, B.J. Grosz, and L.A. Gilliom, "Pronouns, names, and the centering of attention in discourse," *Cognitive Science* 17: 311–347, 1993.

[17] B.J. Grosz and J. Hirschberg, "Some intonational characteristics of discourse structure," *Proc. of the Internat. Conf. on Spoken Language Processing*, Banff, pp. 429–432, 1992.

[18] B.J. Grosz and C.L. Sidner, "Attention, intentions, and the structure of discourse," *Computational Linguistics* 12(3): 175–204, 1986.

[19] M.A.K. Halliday, "Notes on transitivity and theme in English: Part 2," *J. of Ling.* 3: 199–244, 1967.

[8] Much work **has** been done on intonational correlates to discourse in large speech corpora of untrained speakers (e.g. [36, 29]), but this research has in general focused on the prediction of intonational features from text rather than on discourse processing per se.

[20] J. Hirschberg, "Pitch accent in context: Predicting intonational prominence from text," *Artificial Intelligence* 63(1/2): 305–340, 1993.

[21] J. Hirschberg and C.H. Nakatani, "A prosodic analysis of discourse segments in direction-giving monologues," *Assoc. for Comp. Ling.*, 1996.

[22] J. Hirschberg and D.J. Litman, "Now let's talk about *now*: Identifying cue phrases intonationally," *Proc. of the Assoc. for Comp. Ling.*, 1987.

[23] J. Hirschberg and J.B. Pierrehumbert, "The intonational structuring of discourse," *Proc. of the Assoc. for Comp. Ling.*, pp. 136–144, 1986.

[24] S. Hudson-D'Zmura and M.K. Tanenhaus, "Assigning antecedents to ambig. pronouns: The role of the center of attention as the default assignment," In M.A. Walker, et al. (eds.), *Centering Theory in Discourse*, Clarendon Press, pp. 199–226, 1998.

[25] H.-S. Kang, "Acoustic and intonational correlates of the informat. status of referring express. in Seoul Korean," *Lang. & Speech* 39(4): 307–340, 1996.

[26] I. Lehiste, "The phonetic structure of paragraphs," In A. Cohen and S.G. Nooteboom (eds.), *Structure and Process in Speech Perception*, Springer-Verlag, pp. 195–203, 1975.

[27] R.B. Most and E. Saltz, "Information structure in sentences: New information," *Language & Speech* 22: 89–95, 1979.

[28] C.H. Nakatani, "Discourse structural constraints on accent in narrative," In J.P.H. van Santen et al. (eds.), *Progress in Speech Synthesis*, Springer-Verlag, pp. 139–156, 1997.

[29] C.H. Nakatani, *The computational processing of intonational prominence: A functional prosody perspective*, Ph.D. thesis, Harvard University, 1997.

[30] C.H. Nakatani, B.J. Grosz, D.D. Ahn, and J. Hirschberg, "Instructions for annotating discourses," Tech. Rep. TR-21-95, Center for Research in Computing Technology, Harvard, 1995.

[31] S.G. Nooteboom and J.G. Kruyt, "Accents, focus distribution, and the perceived distribution of given and new information," *Journ. of the Acoust. Society of America* 82(5): 1512–1524, 1987.

[32] S.G. Nooteboom and J.M.B. Terken, "What makes speakers omit pitch-accents?," *Phonetica* 39: 317–336, 1982.

[33] J.B. Pierrehumbert and J. Hirschberg, "The meaning of intonation contours in the interpretation of discourse," In P.R. Cohen, et al. (eds.), *Intentions in Communication*, MIT Press, pp. 271–311, 1990.

[34] E.F. Prince, "Toward a taxonomy of given-new information," In P. Cole (ed.), *Radical Pragmatics*, Academic Press, pp. 223–255, 1981.

[35] E.F. Prince, "The ZPG letter: Subjects, definiteness, and information-status," In S. Thompson and W. Mann (eds.), *Discourse Description: Diverse Analyses of a Fund Raising Text*, John Benjamins Publ., pp. 295–325, 1992.

[36] E. Shriberg, R. Bates, A. Stolcke, P. Taylor, D. Jurafsky, K. Ries, N. , R. Martin, M. Meteer, and C. van Ess-Dykema, "Can prosody aid the automatic classification of dialog acts in conversational speech?," *Lang. & Speech* 41(3/4): 443–492, 1998.

[37] K.E.A. Silverman, *The Structure and Processing of Fundamental Frequency Contours*, Ph.D. thesis, University of Cambridge, 1987.

[38] J. Snedeker and J. Trueswell, "Using prosody to avoid ambiguity: Effects of speaker awareness and referential context," *Journ. of Memory and Language* 48: 103–130, 2003.

[39] S. Speer, P. Warren, and A. Schafer, "Intonation and sentence processing," *Proc. of the Internat. Congress of Phonetic Sciences*, Barcelona, 2003.

[40] M. Swerts and R. Geluykens, "Prosody as a marker of information flow in spontaneous discourse," *Language & Speech* 37(1): 21–43, 1994.

[41] M.K. Tanenhaus, M. Spivey-Knowlton, K.M. Eberhard, and J.C. Sedivy, "Integration of visual and linguistic information in spoken language comprehension," *Science* 268: 1632–1634, 1995.

[42] J. Terken, "The distribution of pitch accents in instructions as a function of discourse structure," *Language & Speech* 27(3): 269–289, 1984.

[43] J. Terken and J. Hirschberg, "Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position," *Language & Speech* 37(2): 125–145, 1994.

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

[45] J.J. Venditti, *Discourse Structure and Attentional Salience Effects on Japanese Intonation*, Ph.D. thesis, Ohio State University, 2000.

[46] J.J. Venditti, M. Stone, P. Nanda, and P. Tepper, "Discourse constraints on the interpretation of nuclear-accented pronouns," *Proc. of Speech Prosody*, Aix-en-Provence, 2002.

[47] J.J. Venditti and M. Swerts, "Intonational cues to discourse structure in Japanese," *Proc. of the Internat. Conf. on Spoken Language Processing (ICSLP)*, Philadelphia, pp. 725–728, 1996.

[48] J.J. Venditti, J. Trueswell, M. Stone, and K. Nautiyal, "On-line accented pronoun interpretation in discourse context," Paper presented at the CUNY Conf. on Human Sentence Processing, MIT, 2003.

[49] G. Yule, "Speakers' topics and major paratones," *Lingua* 52: 33–47, 1980.