# Discovering the Sounds of Discourse Structure[*]
## Extended Abstract

### Barbara J. Grosz
Division of Engineering and Applied Sciences

Harvard University

33 Oxford Street

Cambridge, MA 02138 USA

grosz@eecs.harvard.edu

It is widely accepted that discourses are composed of segments and that the recognition of segment boundaries is essential to a determination of discourse meaning (Grosz and Sidner, 1986). Written language has orthographic cues such as section headings, paragraph boundaries, and punctuation which can assist in identifying discourse structure. In spoken language, intonational variation provides essential information about discourse structure. For instance, it may be used to mark structural features of discourse at the global level, such as segment boundaries. Intonation also provides more local information about relations among utterances within a segment, for example indicating whether phrases are parenthetical. It can also help distinguish between different interpretations of phrases that can function either as cue phrases that indicate discourse segment boundaries or sententially to convey domain information. Finally, variations in intonational prominence may be used to convey information about the discourse status of entities referred to by definite noun phrases and pronouns.

An understanding of intonational variation and the ways in which it carries information about discourse characteristics of spoken language is important for computer-based interpretation and generation of speech. From the interpretation perspective, this understanding may provide new techniques for identifying discourse structure. From the generation perspective, it would lead to more natural synthetic speech, making it possible to produce computer speech that is easier for people to understand and less susceptible to misinterpretation.

Three major challenges have faced researchers attempting to discover the relationship between intonational features and the structure of spoken discourse. First, the collection of corpora of spontaneous speech has required the development of new experimental methodologies. Whereas it is straightforward to have the same text read by many speakers, it is much more difficult to obtain similar samples of spontaneous speech from multiple speakers. Second, techniques must be developed to obtain reliable segmentations and labelings of the corpora. Because discourse structure is rooted in semantics rather than syntax, this has proved more difficult than tagging corpora for sentence structure. Third, measures of agreement among segmentations must be designed. In this area too, the semantic nature of discourse structure leads to a more complex problem than comparing sentence parse structures.

This talk will begin with a summary of pilot studies that demonstrated reliable correlations of discourse structure and intonational features (Grosz and Hirschberg, 1992; Hirschberg and Grosz, 1992; Hirschberg and Grosz, 1994). It will then focus on a new corpus of direction-giving monologues, the Boston Directions Corpus (Nakatani et al., 1995a; Hirschberg and Nakatani, 1996). I will describe the methodology we developed to elicit fluent spontaneous direction-giving monologues ranging over a spectrum of planning complexity. Next I will describe the development of annotation instructions used to train labelers to segment spoken discourses (Nakatani et al., 1995b) and will discuss agreement among segmentations on the Boston Directions Corpus obtained using these instructions. Then I will describe results of our analyses of the correlation between discourse structure and intonational features. Finally, I will present a list of challenges for future research in this area.

# References

Barbara Grosz and Julia Hirschberg. 1992. Some intonational characteristics of discourse structure. In John Ohala et al., editor, *Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP-92)*, pages 429-432, Edmonton, Canada. Personal Publishing Ltd.

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

Julia Hirschberg and Barbara Grosz. 1992. Intonational features of local and global discourse structure. In *Proceedings of the Speech and Natural Language Workshop*, pages 441–446. Defense Advanced Research Projects Agency, February.

Julia Hirschberg and Barbara Grosz. 1994. Intonation and discourse structure in spontaneous and read direction-giving. In *Proceedings of the International Symposium on Prosody*, pages 103–109. Japan Society for the Promotion of Science.

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

Christine Nakatani, Julia Hirschberg, and Barbara Grosz. 1995a. Discourse structure in spoken language: Studies on speech corpora. In *Working Notes of the AAAI-95 Spring Symposium on Empirical Methods in Discourse Interpretation*, pages 106–112, Menlo Park, CA. American Association for Artificial Intelligence.

Christine H. Nakatani, Barbara J. Grosz, David D. Ahn, and Julia Hirschberg. 1995b. Instructions for annotating discourse. Technical Report TR-21-95, Harvard University.