

Prosodic Correlates of Discourse Structure

Joseph Tyler

Keywords: *prosody, discourse structure, SDRT*

This production study uses Segmented Discourse Representation Theory (SDRT) (Asher & Lascarides, 2003) to identify prosodic correlates of discourse structure. This study will explore three potential ways discourse structure could correlate with prosody, which I call *relational*, *linear* and *hierarchical*. Pilot data appear to support two of the three. The experimental design involves having subjects read aloud a newspaper article drawn from the SDRT-annotated database DISCOR (Baldrige, Asher, & Hunter, 2007). They are instructed first to study the text's organization and structure, then to paraphrase the text out loud, and finally to read aloud the full article.

In an SDRT analysis, a discourse is segmented and rhetorical relations are coded that hold between those segments. All relations are classified as either coordinating (e.g. CONTRAST) or subordinating (e.g. ELABORATION); coordination and subordination capture the intuition that segments can be at the same or different levels, respectively.

While theoretical work on SDRT has often used constructed examples, the DISCOR corpus applies the same annotation scheme to the natural language text of Wall Street Journal news articles. The following example comes from a 1993 article titled *Murdoch's interest in New York Post seems to pose dilemma for News Corp.*:

(1) 21. *But at issue is whether Mr. Murdoch can overcome federal regulations that would prevent him from bidding for the Post.*

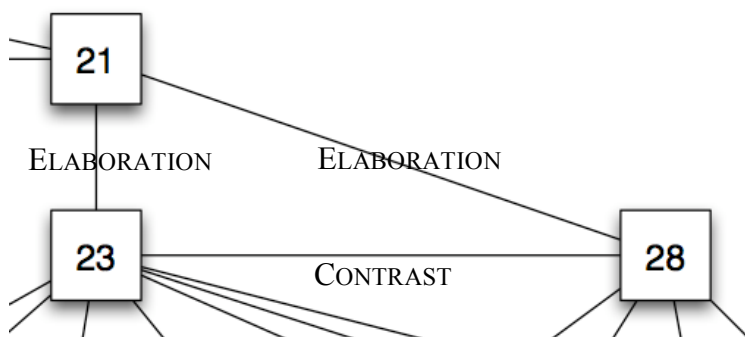
...

23. *Mr. Murdoch is barred from owning both a newspaper and television station in the same market.*

...

28. *Mr. Murdoch could be granted a waiver from the FCC that would allow him to own the Post and his Fox Broadcasting Co. network affiliate.*

In the excerpt above, 21, 23 and 28 are segments of the larger discourse. Segment 21 introduces "federal regulations," about which segments 23 and 28 give more information. And while both 23 and 28 give more information about 21, the information of 28 contrasts with the contribution of 23. In SDRT terms, there are claimed to be relations of the form ELABORATION(21,[23,28]) and CONTRAST(23,28), indicating that segments 23 and 28 together elaborate 21, and that 28 contrasts with 23. Represented graphically:



The elaboration relations are subordinating, indicated by placing the subordinated segments at a lower level in the graph. Contrast relations are coordinating, indicated by relating two segments at the same level.

The *relational* structure approach captures the contrast between coordinating and subordinating relations; the hypothesis is that the ratio of the prosodic measures of the two arguments of a relation will be different for subordinating and coordinating relations. Subordinating relations, e.g. ELABORATION, are hypothesized to have a greater ratio, indicating that the prosodic measure associated with the first segment is greater relative to the second segment; coordinating relations, e.g. CONTRAST, are hypothesized to have a ratio closer to 1, indicating both arguments have relatively similar prosody. With respect to example (1) above, the hypothesis is that the ratio of the prosodic measures of 21 and 23 (related via a subordinating relation) would be different from 23 and 28 (related via a coordinating relation).

Assembling all of a discourse's segments and relations together results in one full-text representation. *Linear* structure is calculated from that full-text representation as the number of nodes intervening between linear segments, while *hierarchical* structure is the number of times a segment is subordinated from the text's topmost segment. Both greater linear strength and height in the hierarchical structure are hypothesized to correlate with wider pitch span, higher f0max, and longer pause duration.

Pause duration is the silence between discourse segments, pitch span is the distance in Hz between a segment's f0max and f0min, while pitch range is simply f0max. While prior studies have only measured f0max, this study will include pitch span because other research has shown f0min *does* vary locally (Ladd & Terken, 1995; Shriberg, Ladd, Terken, & Stolcke, 1996), and relevant discourse structure may be locally variable as well.

In addition to testing these hypotheses, this study seeks to replicate the findings of related prior studies (den Ouden, Noordman, & Terken, 2009; Hirschberg & Grosz, 1992), but with SDRT, and to lay the groundwork for subsequent discourse comprehension studies.

References:

- Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. xxii+526pp, Cambridge, UK: Cambridge U Press.
- Asher, N., & Vieu, L. (2005). Subordinating and Coordinating Discourse Relations. *Lingua*, 115(4), 591-610.
- Baldrige, J., Asher, N., & Hunter, J. (2007). Annotation for and Robust Parsing of Discourse Structure on Unrestricted Texts. *Zeitschrift für Sprachwissenschaft*, 26, 213-239.
- Ladd, D. R., & Terken, J. M. B. (1995). *Modeling pitch range variation within and across speakers: predicting f0 targets when "speaking up"*. Paper presented at the International Congress of Phonetic Sciences, Stockholm.
- Shriberg, E., Ladd, D. R., Terken, J., & Stolcke, A. (1996). *Modeling Pitch Range Variation Within And Across Speakers: Predicting F₀ Targets When "SPEAKING UP" (1996) [1 citations — 0 self]*. Paper presented at the Proceedings of the fourth International Conference on Speech and Language Processing, Philadelphia.