1 Introduction

I want to tell a story about computational approaches to discourse structure. Like all such stories, it takes some liberty with actual events and times, but I think stories put things into perspective, and make it easier to understand where we are and how we might progress.

Part 1 of the story (Section 2) is the past. Here we see early computational work on discourse structure aiming to assign a simple tree structure to a discourse. At issue was what its internal nodes corresponded to. The debate was fierce, and suggestions that other structures might be more appropriate were ignored or subjected to ridicule. The main uses of discourse structure were text generation and summarization, but mostly in small-scale experiments.

Part 2 of the story (Section 3) is the present. We now see different types of discourse structure being recognized, though perhaps not always clearly distinguished. An increasing number of credible efforts are aimed at recognizing these structures automatically, though performance on unrestricted text still resembles that of the early days of robust parsing. Generic applications are also beginning to appear, as researchers recognize the value of these structures to tasks of interest to them.

Part 3 of the story (Section 4) is the future. We now see the need for a mid-line between approaches hostage to theory and empirical approaches free of theory. An empirical approach underpinned by theory will not only motivate sensible back-off strategies in the face of unseen data, but also enable us to understand how the different discourse structures inter-relate and thereby to exploit their mutual recognition. This should allow more challenging applications, such as improving the performance of statistical machine translation (SMT) through the extended locality of discourse structures and the linguistic phenomena they correlate with.

2 Early computational approaches to discourse structure

Early computational work generally assumed discourse structure had an underlying tree structure, similar to the parse tree of a sentence. At issue was what its internal nodes and other formal properties corresponded to. In Rhetorical Structure Theory (Mann and Thompson, 1988), used in both text generation (Scott and de Souza, 1990; Moore, 1995; O’Donnell et al., 2001) and analysis (Marcu, 1996; Marcu, 2000), an internal node corresponded to a rhetorical relation holding between the text units associated with its daughters, and precedence corresponded to their order in the text. In work on generating task instructions (Dale, 1992), each internal node corresponded to the next step to take to accomplish the plan associated with its parent. In (Grosz and Sidner, 1986), which I will return to in Section 4, internal nodes corresponded to speaker intentions, with dominance in the tree corresponding to a daughter node’s intention supporting that of its parent and precedence corresponding to one intention needing to be accomplished before another. The internal nodes in (Moser and Moore, 1996) reflected an attempt to reconcile Grosz and Sidner’s approach with that of Mann and Thompson.

Work that attempted to show that a simple linear model might be a better account for types of expository text (Sibun, 1992) was, by and large, ignored.
3 Current computational approaches to discourse

As well as further elaboration of recursive discourse structures (Asher and Lascarides, 2003; Polanyi et al., 2004), current computational approaches have focussed on discourse structures more easily linked to data: structure associated with changes in topic, structure associated with the function of the parts of a text within a given genre, and structure associated with what one might call higher-order predicate-argument relations or discourse relations.

3.1 Topic structure

Expository text can be viewed as a linear sequence of topically coherent segments (sequences of sentences), where the sequence of topics is either specific to a text or conventionalized (Figure 1).

Interest in topic structure originally came from its perceived potential to improve information retrieval (Hearst, 1994; Hearst, 1997). More recent interest comes from its potential use in segmenting lectures, meetings or other speech events, making them more amenable to search (Galley et al., 2003; Malioutov and Barzilay, 2006).

Computational approaches to topic segmentation all assume that: (1) Relations hold between the topic of discourse segments and the topic of the discourse as a whole (eg, History of Vermont → Vermont). (2) The only relation holding between sister segments, if any, is sequence, though certain sequences may be more common than others (Figure 1). (3) The topic of a segment will differ from those of its adjacent sisters. (Adjacent spans that share a topic will belong to the same segment.) (4) Topic predicts lexical choice, either of all the words of a segment or just of its content words (ie, excluding "stop-words").

Making topic structure explicit (ie, topic segmentation) is based on either semantic-relatedness, where each segment is taken to consist of words more related to each other than to words outside the segment (Hearst, 1994; Hearst, 1997; Choi et al., 2001; Bestgen, 2006; Galley et al., 2003; Malioutov and Barzilay, 2006) or topic models, where each segment is taken to be produced by a distinct, compact lexical distribution (Purver et al., 2006; Eisenstein and Barzilay, 2008; Chen et al., 2009).

3.2 Function-based structure

Texts within a given genre (eg, news reports, errata, scientific papers, letters to the editor, etc.) generally share a similar structure that is independent of topic and reflects the function played by each of its parts. Best known is the inverted pyramid of news reports, consisting of a headline; a lead paragraph, conveying who is involved, what happened, when it happened, where it happened, why it happened, and optionally how it happened; a body that provides more detail; and a tail, containing less important information. This is why the first (ie, lead) paragraph can provide the best extractive summary of a news report.

In the genre of scientific papers (and, more recently, their abstracts), high-level structure comprises the following ordered sections: Objective (also called Introduction, Background, Aim, or Hypothesis); Methods (also called Method, Study Design, or Methodology); Results (also called Outcomes); Discussion and optionally, Conclusions. This does not mean that every sentence within a section realises the same function: Fine-grained functional characterizations of scientific papers (Lialkata et al., 2010; Teufel, 2010) show a range of functions served by the sentences in a section.

Interest in automatic annotation of functional structure comes from its value for summarization (noted above), sentiment analysis, where words may have an objective sense in one section and a subjective sense in another (Taboada et al., 2009), and citation analysis, where a citation may mean different things in different sections (Teufel, 2010).

As with computational models of topic-based structure, computational models of function-based structure make assumptions that may or may not actually hold: (1) Relations hold between the function of a segment and that of the discourse as a whole: While relations may hold between sisters (eg, Methods constrain Results), only sequence has been used in modelling. (2) Function predicts more than lexical choice: it can predict indicative phrases such as “results show” (→ Results) or indicative stop-words such as “then” (→
While the internal structure of a functional segment has usually been ignored in high-level modeling (Chung, 2009; Lin et al., 2006; McKnight and Srinivasan, 2003; Ruch et al., 2007), (Hirohata et al., 2008) found that assuming that properties of the first sentence of a segment differ from those of the rest (as in 'BIO' approaches to Named Entity Recognition) leads to improved performance in segmentation (ie, 94.3% per sentence accuracy vs. 93.3%).

While most functional modelling has been on biomedical text, where texts with explicitly labelled sections serve as “free” training data for segmenting unlabelled texts, there has also been some work on functional segmentation of legal texts and student essays.

3.3 “Higher-order” pred-arg structure

The third type of discourse structure receiving significant attention from the computational world is what can be called higher-order predicate-argument structure, or structure associated with discourse relations. Whereas at the sentence level, pred-arg structures are usually headed by a verb (Kingsbury and Palmer, 2002) or a noun (Gerber et al., 2009), predicate-argument structures in discourse are usually headed by a discourse connective — eg, a conjunction like because or but, or a discourse adverbial like nevertheless or instead.

And just as pred-arg relations within a sentence can conveyed through adjacency (eg, English noun-noun modifiers such as container ship crane operator courses — courses to train operators of cranes that load/unload ships whose cargo is packed in containers), pred-arg relations in discourse can be conveyed through adjacency between clauses or sentences.

The Penn Discourse TreeBank is currently the largest resource manually annotated for discourse connectives, their arguments, and the senses they convey (Prasad et al., 2008). Related resources are also being created for Modern Standard Arabic (Al-Saïf and Markert, 2010), Chinese (Xue, 2005), Czech (Mladová et al., 2008), Danish and Italian parallel treebanks (Buch-Kromann and Korzen, 2010), Dutch (van der Vliet et al., 2011), German (Stede, 2004; Stede, 2008), Hindi (Oza et al., 2009), and Turkish (Zeyrek et al., 2010).

The potential value of being able to automatically recognize these discourse relations, their arguments and their senses comes from their help in question generation (Manhem et al., 2010), extractive summarization (Louis et al., 2010) and sentiment detection (Taboada et al., 2009). So efforts are increasing to automatically recognize them (Elwell and Baldridge, 2008; Lin et al., 2010; Pitler et al., 2008; Pitler et al., 2009; Pitler and Nenkova, 2009; Prasad et al., 2010; Wellner and Pustejovsky, 2007; Wellner, 2008).

4 Future computational approaches to discourse

This story closes with some speculations about the future. I have sketched a past in which computational approaches to discourse structure were hostage to theory and a present in which they are essentially free of theory. What we really want is an empirical approach underpinned by theory, that allows us to understand (at the very least) the ways in which the various types of discourse structures fit together. Early on, (Grosz and Sidner, 1986) attempted to meld a theory of intention-
Discourse Structures and Language Technologies

based discourse structure with a theory of attentional structure (ie, what the conversational participants were attending to), but the link between theory and data was not sufficiently robust. Later attempts to link multiple discourse structures were motivated by purely practical concerns. (Marcu, 2000) used semantic-relatedness methods from topic segmentation to decide what RST-relation to assign to adjacent non-elementary text spans because he could find no other way to do so reliably. (Schilder, 2002) just assumed that RST-relations could only be computed reliably for elementary spans (ie, single clauses or sentences), and used semantic-relatedness methods for other decisions. More recently, (Louis et al., 2010) have shown that features based on RST text structures complement those from discourse relations when it comes to choosing sentences for extractive summaries that are similar to those chosen manually.

While these purely practical links between discourse structures clearly lead to better performance in applications, extensive improvements can, I think, only come with a more theoretically-grounded understanding of how the different types of discourse structure fit together.

References


Emily Pitler, Annie Louis, and Ani Nenkova. 2009. Automatic sense prediction for implicit discourse relations in text. In *Proc. 47th Meeting of the ACL and 4th Int’l Joint Conference on Natural Language Processing*.


