UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A Theory of Discourse Structure

Permalink

Journal

Author

Grosz, Barbara J.

Publication Date

1986

Peer reviewed

A THEORY OF DISCOURSE STRUCTURE

Barbara J. Grosz
Artificial Intelligence Center, and
Center for the Study of Language and Information
SRI International

ABSTRACT

1. Overview

In this talk I will present the basic elements of a computational theory of discourse structure.[1] A proper account of discourse structure is needed both as the basis of an account of discourse meaning (a semantic task) and to underlie a model of discourse processing. It provides the former by specifying the basic units a discourse comprises and the ways in which they can relate. It plays a key role in discourse processing by stipulating constraints on those portions of a discourse to which any given utterance in the discourse must be related.

An account of discourse structure is closely related to two questions: What individuates a discourse? What makes a discourse coherent? That is, faced with a sequence of utterances, how does one know whether they constitute a single discourse, several (perhaps interleaved) discourses, or none? Likewise, how does a discourse participant know where and how an utterance belongs in a discourse.

In the theory, discourse structure is intimately connected with two nonlinguistic notions: intention and attention. Intentions play a primary role in explaining discourse structure, defining discourse coherence, and providing a coherent conceptualization of the term "discourse" itself. Attention is an essential factor in explicating the processing of utterances in discourse.

A major claim of the theory that Sidner and I are developing [Grosz and Sidner 86] is

---

[1]Substantial portions of this abstract are taken from a paper coauthored by C.L. Sidner [Grosz and Sidner 86]. The theory I describe is being developed jointly with her. More details may be found in the paper.

that the structure of any discourse is a composite of three distinct but interacting components: (1) the structure of the actual sequence of utterances in the discourse; (2) a structure of intentions; (3) an attentional state. In examining attentional state, I will show that it has at least two separate constituents: one corresponds to the *global* focus of attention of the participants in a discourse and the other to their more local *centering* of attention at any given utterance in the discourse.

The distinction among the three components is essential to an explanation of various linguistic phenomena, including interruptions, the differential use of certain types of referring expressions, and the use of certain phrases that at times function solely to affect discourse segmentation and structure. Examples of each of these phenomena and the role of the components in explaining them will be given in the talk.

By providing an overall framework within which to answer questions about the relevance of various segments of discourse to one another and to the overall purposes of the discourse participants, the theory has implications for natural-language processing work in general. In particular, various properties of the intentional component suggest problems with approaches to discourse coherence based on selecting discourse relationships from a fixed set of alternative rhetorical patterns (e.g., [Hobbs 79], [Mann 83], [Reichman 81]) as well as indicating several issues that must be confronted in expanding speech-act-related theories (e.g., [Allen 80], [Cohen 80], [Allen 83]) from coverage of individual utterances to coverage of extended sequences of utterances in discourse.

Most of this presentation will be concerned with specifying an abstract model of discourse structure; in particular, the definitions of the components will abstract away from the details of the discourse participants. Both the construction of a computer system that can participate in a discourse (i.e., one that **is** a language user) and the specification of a psychological model of language use require the appropriate projection of this abstract model onto properties of a language user and specification of additional details; for example, memory for linguistic structure must be specified and means for encoding attentional state, and appropriate representations of intentional structure must

be provided). Although I will not address such issues completely, I will examine certain key processing questions.

2. Linguistic Structure

Utterances--the actual saying or writing of particular sequences of phrases and clauses-- are the basic elements of the linguistic structure. Just as the words in a single sentence form constituent phrases, the utterances in a discourse are naturally aggregated into *discourse segments*. The linguistic structure consists of these discourse segments and embedding relationships among them.

The utterances in a segment, like the words in a phrase, serve particular roles with respect to that segment. In addition, the discourse segments, like the phrases, fulfill certain functions with respect to the overall discourse. Although two consecutive utterances may be in the same discourse segment, it is also common for two consecutive utterances to be in different segments. It is also possible for two utterances that are nonconsecutive to be in the same segment. An individual segment may include a combination of subsegments and utterances only in that segment (and not members of any of its embedded subsegments).

One claim of the theory is that the embedding relationships are a surface reflection of relationships among elements of the intentional structure. However, various elements of the linguistic structure are crucial in conveying information about the intentional structure so that these two components are mutually constraining. Because the linguistic structure is not strictly decompositional, various properties of the discourse (most notably the intentional structure) are functions of properties of individual utterances and properties of segments. In addition to giving examples that illustrate the nonstrict decompositionality of discourse, I will discuss problems that arise in determining the intentional structure because of it.

There is a two-way interaction between the discourse segment structure and the utterances constituting the discourse: linguistic expressions can be used to convey information about the discourse structure; conversely, the discourse structure constrains

the interpretation of expressions (and hence affects what a speaker says and how a hearer will interpret what is said). Not surprisingly, linguistic expressions are among the primary indicators of discourse segment boundaries. The explicit use of certain words and phrases (e.g., "in the first place") and more subtle clues, such as changes in tense and aspect, are included in the repertoire of linguistic devices that function, wholly or in part, to indicate these boundaries [Reichman-Adar 84, Cohen 83, Polanyi 83]. I will give examples that show these linguistic boundary markers can be divided according to whether they explicitly indicate changes in the intentional structure or in the attentional state of the discourse. The differential use of these linguistic markers provides one piece of evidence for considering these two components to be distinct.

Because these linguistic devices function explicitly as indicators of discourse structure, it becomes clear that they are best seen as providing information at the discourse level, and not at that of the sentence; hence, certain kinds of questions (e.g., about their contribution to the truth conditions of an individual sentence) do not make sense. For example, in the utterance "Incidentally, Jane swims every day," the "incidentally" indicates an interruption of the main flow of discourse rather than affecting in anyway the meaning of "Jane swims every day." Jane's swimming every day could hardly be fortuitous.

Just as linguistic devices affect structure, so the discourse segmentation affects the interpretation of linguistic expressions in a discourse. Referring expressions provide the primary example of this effect. The segmentation of discourse constrains the use of referring expressions by delineating certain points at which there is a significant change in what entities (objects, properties, or relations) are being discussed. For example, there are different constraints on the use of pronouns and reduced definite-noun phrases within a segment than across segment boundaries. While discourse segmentation is obviously not the only factor governing the use of referring expressions, it is an important one.

3. Intentional Structure

A rather straightforward property of discourses, namely, that they (or, more accurately, those who participate in them) have an overall purpose, turns out to play a fundamental role in the theory of discourse structure. Thus, intentions of a particular sort and a small number of relationships between them provide the basic elements of the intentional structure.

Although typically the participants in a discourse may have more than one aim in participating in the discourse (e.g., a story may entertain its listeners as well as describe an event; an argument may establish a person's brilliance as well as convince someone that a claim or allegation is true), we distinguish one of these purposes as foundational to the discourse. We will refer to it as the *discourse purpose (DP)*. From an intuitive perspective, the discourse purpose is the intention that underlies engaging in the particular discourse. This intention provides both the reason a discourse (a linguistic act), rather than some other action, is being performed and the reason the particular content of this discourse is being conveyed rather than some other information. For each of the discourse segments, we can also single out one intention--the *discourse segment purpose (DSP)*. From an intuitive standpoint, the DSP specifies how this segment contributes to achieving the overall discourse purpose.

Typically the initiator of a discourse (e.g., the writer of a text or the first speaker in a dialogue) will have a number of different kinds of intentions that lead to initiating a discourse. One kind might include intentions to speak in a certain language or to utter certain words. Another might include intentions to amuse or to impress. The kinds of intentions that can serve as discourse purposes or discourse segment purposes are distinguished from other intentions by the fact that they are intended to be recognized (cf. [Allen 80], [Sidner 85]), whereas other intentions are private; that is, the recognition of the DP or DSP is *essential* to its achieving its intended effect. Discourse purposes and discourse segment purposes share this property with certain utterance-level intentions that Grice [Grice 69].

It is important to distinguish intentions that are intended to be recognized from other

kinds of intentions that are associated with discourse. Intentions that are intended to be recognized achieve their intended effect only if the intention is recognized. For example, a compliment achieves its intended effect only if the intention to compliment is recognized; in contrast, a scream of "boo" typically achieves its intended effect (scaring the hearer) without the hearer having to recognize the speaker's intention.

We have identified two structural relations that play an important role in discourse structure: *dominance* and *satisfaction-precedence*. An action that satisfies one intention, say DSP1, may be intended to provide part of the satisfaction of another, say DSP2. When this is the case, we will say that DSP1 *contributes to* DSP2; conversely, we will say that DSP2 *dominates* DSP1 (or DSP2 *DOM* DSP1). The dominance relation invokes a partial ordering on DSPs that we will refer to as the *dominance hierarchy*. For some discourses, including task-oriented ones, the order in which the DSPs are satisfied may be significant, as well as being intended to be recognized. We will say that DSP1 *satisfaction-precedes* DSP2 (or, DSP1 *SP* DSP2) whenever DSP1 must be satisfied before DSP2.[2]

The range of intentions that can serve as discourse, or discourse segment, purposes is open-ended (cf. [Wittgenstein 53], paragraph 23), much like the range of intentions that underlie more general purposeful action. There is no finite list of discourse purposes, as there is, say, of syntactic categories. It remains an unresolved research question whether there is a finite description of the open-ended set of such intentions. However, even if there were finite descriptions, there would still be no finite list of intentions from which to choose. Thus, a theory of discourse structure cannot depend on choosing the DP/DSPs from a fixed list (cf. [Reichman-Adar 84], [Schank 82], [Mann 83]), nor on the particulars of individual intentions. Although the particulars of individual intentions, like a wide range of common sense knowledge, are crucial to understanding any discourse, such particulars cannot serve as the basis for *determining* discourse structure.

---

[2]These two relations are similar to ones that play a role in parsing at the sentence level: immediate dominance and linear precedence. However, the dominance relation, like the one in Marcus and Hindle's D-theory [Marcus et al. 83], is partial (i.e., nonimmediate).

What is essential for discourse structure is that such intentions bear certain kinds of structural relationships to one another. Since the CPs can never know the whole set of intentions that might serve as DP/DSPs, what they must recognize is the relevant structural relationships among intentions. Although there is an infinite number of intentions, there are only a small number of relations relevant to discourse structure that can hold between them.

4. Attentional State

The third component of discourse structure, the attentional state, is an abstraction of the participants' focus of attention as their discourse unfolds. Attentional state contains information about the objects, properties, relations, and discourse intentions that are most salient at any given point. It essentially summarizes information from previous utterances crucial for processing subsequent ones thus obviating the need for keeping a complete history of the discourse. It is inherently dynamic, recording the objects, properties, and relations that are salient at each point in the discourse. Thus attentional state serves an additional purpose: namely, it furnishes the means for actually using the information in the other two structures in generating and interpreting individual utterances.

In this theory, the global component of attentional state is modelled by a set of *focus spaces*; changes in attentional state are modelled by a set of transition rules that specify the conditions for adding and deleting spaces. We call the collection of focus spaces available at any one time the *focusing structure* and the process of manipulating spaces *focusing*.

The global focusing process associates a focus space with each discourse segment; this space contains those entities that are salient--either because they have been mentioned explicitly in the segment or because they became salient in the process of producing or comprehending the utterances in the segment. The focus space also includes the DSP; the inclusion of the purpose reflects the fact that the CPs are focused not only on what they are talking about, but also on why they are talking about it. Note however, that although each focus space contains a DSP, the focus structure does *not* include the

intentional structure as a whole.

The local component of attentional state is modelled by assigning to each utterance a *backward-looking center* and a set of *forward-looking centers* [Grosz et al. 83, Grosz et al. 86]. Centering rules provide constraints on how various entities can be realized in an utterance (e.g., what surface forms can be used to refer to a given object). I will give examples that illustrate differences in acceptability according to whether these rules are followed or not.

It is important to note that attentional state component is not equivalent to cognitive state, but is only one of its components. Cognitive state is a richer structure, one that includes at least the knowledge, beliefs, desires, and intentions of an agent, as well as the cognitive correlates of attentional state as modelled in this paper.

Acknowledgements

References

[Allen 80] Allen, J.F., and Perrault, C.R.
Analyzing intention in dialogues.
Artificial Intelligence 15(3):143-178, 1980.

[Allen 83] Allen, J.F.
Recognizing Intentions from Natural Language Utterances.
In M. Brady and R.C. Berwick (editors), *Computational Models of Discourse*, pages 107-166. Massachusetts Institute Technology Press, 1983.

[Cohen 80] Cohen, P.R. and Levesque, H.L.
Speech Acts and the Recognition of Shared Plans.
In *Proc. of the Third Biennial Conference*, pages 263-271. Canadian Society for Computational Studies of Intelligence, Canadian Society for Computational Studies of Intelligence, Victoria, B. C., May, 1980.

[Cohen 83] Cohen, R.
A Computational Model for the Analysis of Arguments.
Technical Report CSRG-151, Computer Systems Research Group, University of Toronto, October, 1983.

[Grice 69] Grice, H.P.
Utterer's Meaning and Intentions.
Philosophical Review 68(2):147-177, 1969.

[Grosz and Sidner 86]
Grosz, B.J. and Sidner, C.L.
Attention, Intentions, and the Structure of Discourse.
Computational Linguistics 12, 1986.

[Grosz et al. 83] Grosz, B.J., Joshi, A.K., Weinstein, S.
Providing a Unified Account of Definite Noun Phrases in Discourse.
In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, June, 1983.

[Grosz et al. 86] Grosz, B.J., Joshi, A.K., and Weinstein, S.
Towards a Computational Theory of Discourse Interpretation.
, 1986.
in preparation.

[Hobbs 79] Hobbs, J.
Coherence and Co-references.
Cognitive Science 3(1):67-82, 1979.

[Mann 83]   W.C. Mann and S.A. Thompson.
Relational Propositions in Discourse.
Technical Report RR-83-115, Information Sciences Institute, November, 1983.

[Marcus et al. 83]
Marcus, M. P., Hindel, D., and Fleck, M. M.
D-Theory: Talking about Talking about Trees.
In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 129-136. The Association for Computatinal Linguistics, Cambridge, MA, June, 1983.

[Polanyi 83]   Polanyi, L. and Scha, R.
On the Recursive Structure of Discourse.
In K. Ehlich and H. van Riemsdijk (editors), *Connectedness in Sentence, Discourse and Text*, pages 141-178. Tilburg University, 1983.

[Reichman 81]   R. Reichman.
Plain-speaking: A theory and grammar of spontaneous discourse.
PhD thesis, Department of Computer Science, Harvard University, 1981.
Also, BBN Report No. 4681, Bolt Beranek and Newman Inc., Cambridge, MA.

[Reichman-Adar 84]
Reichman-Adar, R.
Extended Person-Machine Interface.
Artificial Intelligence 22(2):157-218, March, 1984.

[Schank 82]   Schank, R.C., Collins, G.C., Davis, E., Johnson, P.N., Lytinen, S., Reiser, B.J.
What's the Point?
Cognitive Science 6(3):255-275, July-September, 1982.

[Sidner 85]   Sidner, C.L.
Plan parsing for intended response recognition in discourse.
Computational Intelligence 1(1):1-10, February, 1985.

[Wittgenstein 53] Wittgenstein,L.
Philosophical Investigations.
Oxford Press, 1953.