Springer
Berlin
Heidelberg
New York
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

Computational and Conversational Discourse

Burning Issues – An Interdisciplinary Account

Edited by

Eduard H. Hovy

Information Sciences Institute, University of Southern California
4676 Admiralty Way, Marina del Rey, California 90292-6695, USA

Donia R. Scott

Information Technology Research Institute, University of Brighton
Lewes Road, Brighton BN2 4NT, UK

Springer
Published in cooperation with NATO Scientific Affairs Division

2.2 Discourse Representation Theory

Discourse Representation Theory (Kamp 1981, Kamp and Reyle 1993, Asher 1993, Asher and Kamp 1995) aims to account for the compositional semantic interpretation of entire discourses and to unify its account of truth conditions for discourses with an account of the mental representation of the discourse built by the speaker/hearer. The discourse representation structure contains separate symbols for each of the distinct events and other abstract types introduced into the discourse. In our approach discourse structure operates over sets of these abstract types. These sets (discourse segments) form internally coherent eventuality descriptions or propositional content. Coherence and segmentation inferences operate over these abstract types to produce mental models of coherent discourse. The relata are individual event reference markers, or sums of events (cf. Section 4.1). Coherence predicates are added to the formal discourse representation structure of a text, to form a cognitive model. This is done bottom-up: as each new sentence is interpreted, coherence and segmentation inferences are attempted, and the discourse structure is recovered and modified sentence by sentence.

Our local coherence relation assignment algorithm assigns a coherence relation over two events (or other abstract types) in a cognitive model (Dahlgren, 1989). The algorithm was developed by examining signals of coherence relatedness in a corpus. The information used by the local algorithm includes syntactic properties of the source clause, connectives in either the source or target, the temporal order of the events in the source and target, naive semantic information (world knowledge) associated with the verbs, semantic information such as types of adverbials, mood, and agentiveness. The algorithm was hand tested in the original corpus with 97.5% accuracy.

2.3 Naive Semantics

In order to draw implicit discourses inferences in the absence of overt markers, people employ world knowledge. We briefly describe our lexical theory which incorporates commonsense knowledge to substantiate the feasibility of our approach to coherence. Naive Semantics (NS) (Dahlgren and McDowell, 1986, Dahlgren 1988a, 1991, 1994) is a theory of word sense meaning representation in which associated with each sense of each content word (noun, verb, adjective) is a naive theory of the sort of object, action or property named by the word. The propositions in word sense meanings (e.g., "Cats have tails", "The goal of investing is making a profit") contribute non-monotonically to the truth conditions of sentences of which they become a part. Though not "true", they are and must be close enough to true, enough of the time, for people to refer correctly to real objects and events, and to communicate using language. In principle, the number of feature values could be as large as the number of words in English (Schubert et al. 1979) and thus they do not form a set of primitive concepts from which all

others are derived. This open-endedness is countervailed by the shallowness of lexical knowledge (Dahlgren, 1991). Naive semantic representations consist only of that shallow knowledge which people must deploy in order to parse and interpret the text, and to find the antecedents of anaphoric expressions. While the amount of world knowledge associated with a word sense varies with the expertise of the individual, we have found that only a few propositions are required in order to interpret and disambiguate text in many domains, including finance, popular medicine, law, intelligence and software. We have found that the same shallow lexical knowledge which informs structural and word sense disambiguation is also sufficient for many other forms of reasoning: anaphora resolution (Wada, 1994), coherence (Dahlgren 1989) and discourse segmentation, and relevance reasoning (Dahlgren, 1994). With independently-motivated naive semantic representations, there is sufficient world knowledge to assign coherence relations for all clauses in Study 1's corpus.

In particular, for verbs, Naive Semantics incorporates the finding that people conceive of actions in terms of their causal implications (Graesser and Clark 1985b). Naive Semantic representations of verb senses consist of implications such as causes, goals, enablements, consequences, constituent events and states, and typical follow-on events and states, as well as selectional restrictions. Thus, the psycholinguistically-motivated features in verb representations form the basis of generating coherence inferences. This is no accident, because they are derived from studies of human text understanding. For the verb "invest", the goal is to make a profit, the enablement is having money, and so on. Thus the coherence relation between the investing and profiting events in (1) is derived by inspecting the naive semantic lexicon.

2.4 Other Approaches to Coherence

In comparison with other work, our approach is similar to that of van Dijk and Kintsch (1983). They define coherence by whether sentences in a discourse describe related facts in some possible world, and they assume that large amounts of world knowledge are employed in building a cognitive model of a discourse. We differ from them in defining coherence as relating discourse events, rather than as relating sentences. Furthermore, we clarify the question of truth conditions as opposed to naive (or heuristic) inference regarding discourse interpretation. And we provide an algorithm. We draw upon Hobbs (1979) and Mann and Thompson (1987) for coherence relations. However, we define them as relating discourse events (and other abstract types) and structures of these in a cognitive event model, rather than as relating utterances, clauses, or spans of discourse. The cognitive representation of an event model, after parsing and formal semantic interpretation is complete, is made to cohere and have structure. For them, coherence is essentially a property of presentation style or of the speaker's intended effect on the hearer. We agree with Polanyi (1988) that this aspect of coherence belongs in

a level of theory above that of cognitive models of interpretation of discourse. For us, coherence is essentially a property of mental models (Johnson-Laird, 1983) which finds its origin in beliefs about relationships among real events. Our approach appeals to cognitive strategies and beliefs people use all the time, whether thinking verbally or not. In the section "Global Coherence and Segmentation", we carefully scrutinize the details of several of these approaches in relation to the question of tree structure for discourse.

3. Approaches To Discourse Segmentation

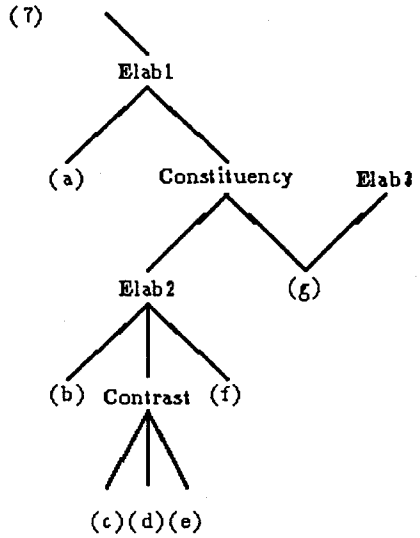
In contrast to many linguists and sociologists (cf. Chapters 2, Martin, and 1, Schegloff), computational researchers tend to the view that discourse structure is a dominance hierarchy, one that is generative (cf. Hobbs, Chapter 6 of this volume). However, beyond these assumptions there is little agreement about the elements in the dominance hierarchy, the meaning of arcs in the hierarchy, and the level of grammar at which discourse structure is built. In this section we examine the main proposals in discourse theory as they address these foundational issues. A new theory will be proposed which builds discourse structure bottom-up, accounts for anaphora resolution constraints, and gives a unified account of dominance in the discourse tree as topic-relatedness. The theory assumes Asher's (1993) account of event summation and abstract type anaphora. The elements in the theory will be justified in comparison with other proposals.

3.1 Hobbs

Hobbs (1979) builds discourse structure into an undirected graph according to coherence relations between clauses. Higher nodes in the structure are binary coherence relations, and the leaves are clauses. The leaves preserve the sequence of the clauses in the text. An example is the text in (6).

(6) (a) I had already, as I told you, learned a little bit about hitchhiking,
(b) I'd split out and two or three times, then come back
(c) The one—my first trip had been to Geneva New York,
(d) and then I'd once or twice gone to—twice I'd gone to California,
(e) And then I'd cut down through the South,
(f) And I had sort of covered the United States.
(g) One very beautiful summer... that I spent in Idaho

Hobbs analyzes this monologue clause by clause, finding a coherence relation between each clause and some other clause. Such a procedure accords with Hobbs' convincing claim that if no relation between a new clause and what has come before can be found, the text seems incoherent. The result is tree (7).



The story in (6) has two main segments, an elaboration of the speaker's hitchhiking experience in clauses (b–f), and an example of one particular trip starting with (g). These segments are readily identified as the subgraphs under ELABORATION 2 and ELABORATION 3 (not shown). Thus the graph captures our intuitive feeling for the structure of the discourse to some extent. However, while the graph preserves linear order and segments are visible, there is no root to the graph, and no semantics of dominance. But in many genres, including some narratives, there is an intended outline to the discourse dominated by a single global discourse topic (Van Dijk and Kintsch 1983, Hinds 1984, Reichman-Adar 1984). This is true for (6), which is clearly about hitchhiking. Thus (a) should be the root of the graph and dominate the structure of the entire narrative. Further, the clause (g) is both an elaboration of clause (a), that is, it describes one hitchhiking event, and is also elaborated by additional clauses (not shown), but it's role relative to both (a) and the additional claues is obscured. In the graph, the node Elab 3 is higher than (g), but (g) is the topic of the remaining (h–m), and should dominate them (Grosz and Sidner 1986, Polanyi 1988).

Another small text, (8), extracted from our *Wall Street Journal* corpus, illustrates the problem with the dominance relation in Hobbs-type discourse structures.

(8) (a) Last fall one of the most promising new shows on television was canceled because of the writer's strike.

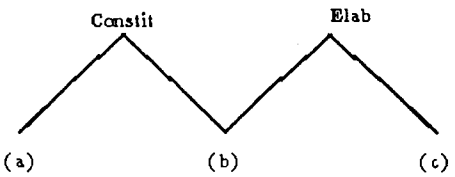
(b) In general network viewing is on the decline.

(c) Viewers are reacting to delays, shoddily produced episodes, and changes in schedules by switching from the networks to cable television and videos.

(d) This will probably cost the networks hundreds of millions in revenues.

Intuitively, this text is about a decline in network viewing, exemplified in (a), elaborated in (c), and given an importance in (d). This structure is not transparent in a discourse tree constructed using the Hobbs method. First clauses (a) and (b) are connected via a CONSTITUENCY relation. Then (b) and (c) are connected by ELABORATION (as with (g) and the CONTRAST node dominating (h–m) in (9)).

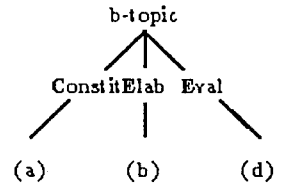
(9)



Now where should EVAL clause (d) be inserted? As in the Hobbs tree of (6), we could cross over, yielding the tree in (10). But this tree is most unsatisfying because our intuitive feeling for the structure of the text (and the way it would be segmented by many people) is the tree in (11).

(10)

(11)



3.2 Our Initial Proposal

In order for the tree to reflect dominance in the discourse structure, so that the higher the node in the tree, the more closely related it is to the global topic (as in VD&K), each node has to be a clause. Then the root can be the topic, nodes one down from the root the main sub-topics in the discourse, and so on.

Clausal Definition of a Discourse Segment

A discourse segment is a possibly discontiguous set of clauses which share a single relation to some other segment or clause. Usually, but not always, segments are contiguous in the discourse surface structure.

We add a new node for each clause and label it with a coherence predicate which relates it to one above. Nodes are clauses and arcs connect them to the clauses with which they most directly cohere in the discourse. Thus, (8) would be represented as in (11). The root of the discourse structure tree becomes the topic of the discourse. Nodes it dominates are those clauses which most directly cohere with it. Those nodes dominate subtrees which elaborate the major sub-topics in the discourse. Higher nodes are more related to the global topic (more salient) in the discourse, subtrees are subsegments, and nodes relate a discourse event or set of events (or other abstract types) mentioned in a dominating node with some event reference marker or set of reference markers introduced at that node. Coherence relations tend to be binary, and this fact is captured by connecting related clauses. In some cases a whole segment relates to a clause.

In some cases, a summary clause introduces or closes the segment. In those cases, the discourse structure is easy to represent. The summary clause can be placed at the node dominating the segment it summarizes. For example, suppose that (8a) is elaborated as in (12).

(12) (a) Last fall "Tattingers" was one of the most promising shows on tv.

(a1) "Tattingers" had one of TV's best producers.

(a2) The star came from another successful series.

(a3) The writers were Emmy winners.

Here the top node of the tree is the generalization (a), and the three elaborations can be attached below it. However, sometimes such subsegments have no introductory or concluding summary. A discourse could start out with examples, and then go on to the topic sentence as in (13).

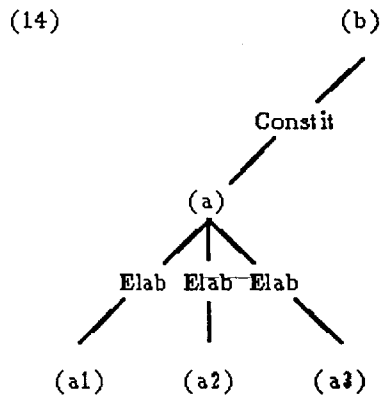
(13) (a1) "Tattingers" was cancelled.

(a2) "The Cosby Show" was postponed.

(a3) Several new series never materialized.

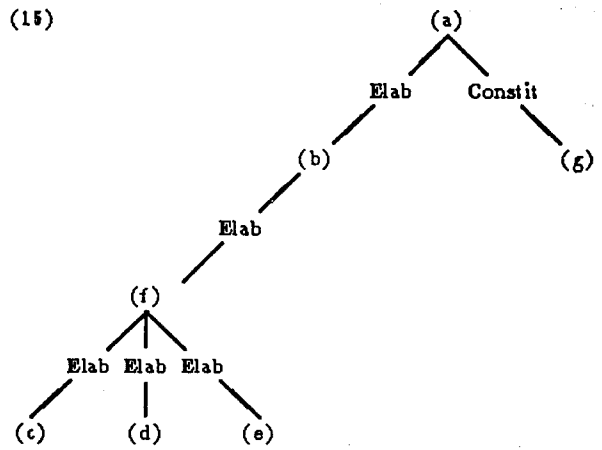
(b) So it goes in this dismal network season.

In this case, following Polanyi (1988) and Asher (1993), we introduce an implicit topic as in (14).



Returning to Hobbs' hitchhiking text in (8), our proposed tree is given in (15).

In (15), the fact that the text is about hitchhiking, and the remainder of the text elaborates this topic, is reflected in the structure of the tree. All nodes are dominated by the topic node (a). Similarly, the elaborations of the generalizations in (b) and (f) are dominated by the nodes which introduce (b) and (f). An arc label is a coherence predicate relating the node below to that above. The nodes are (the content of) clauses. So a coherence predicate elaborate (b,a) can be read off the root. A summary should be higher in the tree, because it is more important and memorable than the details (Van Dijk and Kintsch 1983, Grosz and Sidner 1986). In the proposed method not only the right-hand side of the tree, but the bottom is open for adding new information to the tree. The analysis of clause (f) would indicate that interior nodes of the tree are open for additions. At the point in the processing when (f) is encountered, a whole subtree has already been built under ELABORATION(b,a), (the subtree covering (c,d,e)). This node is modified by inserting (f) to dominate the subtree (c,d,e), and ELABORATION(f,b). The advantage of permitting additions to the interior of the tree is that the final result will look like an outline of the content of discourse. This makes an empirical prediction that the fewer such corrections and the less backtracking required for its interpretation, the more easily the discourse is processed. It also accords with the finding that in memory for text people build an organization reflecting content (Van Dijk and Kintsch 1983, Morrow et al. 1987). The disadvantage is the computational complexity of redrawing the tree as a restructuring clause is encountered. A less complex mode would keep nodes on the right open, as in Polanyi (1988), and account for reorganization and restructuring of mental models as a later stage of processing.



3.3 Polanyi

In Polanyi, the basic discourse unit is the propositional content of a clause, and larger units are built up from these basic units. The nodes are labeled with the propositional content introduced at that node, and lower nodes expand upon the semantic content of higher nodes. If a text implies a generalization that goes unsaid, it is inferred and added to the representation of the discourse. Polanyi assumes that there are only two ways two clauses can be related, as coordinated or subordinated, so that discourse structure trees are binary. As the discourse tree is built up, a new coordinate or subordinate node is added relating a new clause to just one other. In some cases propositions are repeated, and in others they are added to express the content of coordinated propositions. The deep discourse trees it generates do not reflect the intuitive chunks or segments of text which elaborate a topic.

In our theory discourses are broken into segments which are related by coherence predicates. The discourse structure is built up from the expressed content, topic inferences are only added if required by the discourse structure. In contrast, Polanyi's binary parsing method forces topic inferences for every pair of clauses under an subordination node where no such topic is directly expressed. Another problem with Polanyi's theory is that the root of the binary tree does not in general correspond to the global discourse topic, and subtrees to subtopic. Height in the tree is unrelated to topic. In contrast, in our theory the tree structure reflects outlining. The final problem with the Polanyi theory, which holds as well for Hobbs, is that it places clauses at the nodes. This leads to an overly fine-grained discourse structure, and collapses the distinction between syntactically signaled semantic properties (embedding, conditionals, quantifiers) and the

discourse structure which explains the intuitive notions of coherence and outlining. We return to these problems below.

3.4 Mann and Thompson

An approach which builds in the fact that a coherence relation can relate one clause to a set of clauses (or span) is that of Mann and Thompson (1987). Coherence relations relate a clause to another clause or to a span.. We follow Mann and Thompson in relating sets of clauses as well as individual clauses, although for us the relata are contents of clauses rather than the clauses themselves. Our approach differs from theirs in aiming to build discourse theory upon semantic theory. They are interested in diagramming the surface level of discourse, clause-in-the-text by clause-in-the-text. However, if discourse structure belongs in a cognitive model, and is thus accounted for in a semantic theory of content plus naive inferences, surface relations are not the level at which discourse structure can be defined. (as Polanyi, VD&K, Asher, etc., have established). Second, dominance in the diagrams is unrelated to importance in the text, and the topic is not necessarily at the top. Third, the spans must be contiguous in the text, and the leaves of the diagram are the clauses of the text in the order they appear. The discussion of the Hobbs text (6) shows where the discourse structure doesn't always follow exactly the order of clause presentation.

3.5 Grosz and Sidner

Our discourse segmentation method is similar to that of Grosz and Sidner (1986), hereafter G&S. The segmentations result in a hierarchy in which the root is the discourse topic, and subtrees correspond to subtopics. Furthermore, the trees are non-binary. Our treatment of anaphor resolution differs in some respects, but agrees with G&S that discourse structure is an important determinant of the availability of discourse entities as antecedents of anaphors. There are two key differences. First, G&S are concerned with explaining the effect of speaker intentions on the discourse structure, while we are concerned with the recovery of the semantic content of the discourse, and assign intentions and planning to a higher-level theory. This follows Polanyi (1988), where the levels are: discourse structure, genre, speech event and interaction. Speaker intention belongs at the level of speech event, and is one of the determinants of genre. Secondly, G&S abjure coherence relations as under-defined (a problem this paper hopes to address). Our Study 1 of discourse segmentation (described below) shows that coherence relations are fundamental to the recovery of discourse structure, at least in some genres. In newspaper commentary, change in coherence relation is the main signal of change of discourse segment in a commentary genre, while explicit cue phrases, are infrequent. In a corpus of 16000 words, an average of 90% of

segment boundaries change discourse relation, while only 16% have a cue phrase. Coherence is the most informative factor determining discourse segmentation and thus constraining anaphora resolution. Even though the theory of coherence relations is at a formative stage, it is a necessary element of any theory that hopes to account for the facts.

In G&S, the two example texts illustrate the fact that genre is an important element of the theory of discourse. Their examples belong in the rhetoric and task-oriented genres. In the commentary genre we have studied, the topic-dominated discourse structure tree is explained by and determined by coherence reasoning about events. In other genres, the structure is very different. In the rhetoric text, the topic is often a proposition, as in G&S's example, where the topic proposition is "Parents shouldn't allow children to go to movies". The segments dominated by the root are background and arguments justifying the topic proposition in some logical argument form such as syllogism, induction, etc. On the other hand, the pump example is in the task-oriented dialogue genre, where the discourse structure is determined by the structure of a task (and even, to some extent, by the structure of the artifact being built), rather than by speaker hypotheses concerning the coherence of events.

3.6 Our Revised Proposal

Thus, the revised version of discourse structure theory we propose is as follows:

> The propositional content of clauses is placed at nodes. Each node is labeled with a coherence relation involving the content at that node, and a higher node. Thus arcs in the tree signify that the nodes they connect are coherence related. Height in the tree corresponds to topic-relatedness in some genres. The deeper the tree, the more complex the outline, and the longer the discourse. Subtrees are segments of the discourse. The global topic of the discourse is at the root of the tree. Interruptions and repairs are not part of the discourse structure tree.

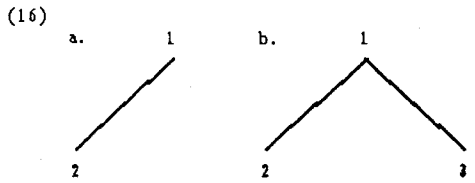
Below, we modify this theory to make trees shallower by recognizing event summation in cognitive representations.

3.7 Study 1 of Naive Semantics and Discourse Segmentation

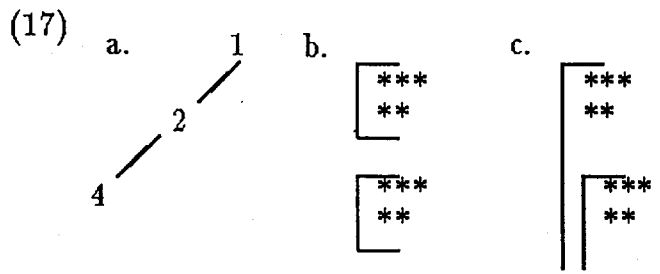
The purpose of the study was to discover the signals of discourse segment boundaries. The study was empirical in that a corpus of Wall Street Journal commentary was analyzed and segmented, the potential segmenting factors were assigned for all clauses (those at segment boundaries and those not), and the information content level of each factor was tabulated. Coherence relations and segmentation were intuited. A check on segmentation boundaries was made by

having two judges segment independently. However, as pointed out in Chapter 7 (Passonneau and Litman), considerable variability exists in segmentation.

A discourse segment for purposes of this study is a chunk of a discourse which has the same outermost coherence relation to some other discourse entity or segment. In other words, clauses deeply embedded inside a subtopic share a coherence relation to the topic with the dominating subtopic node and all of its daughters. They have a unified function in relation to the rest of the discourse. Two judges (the author and Carol Lord) segmented 16000 words of WSJ commentary using the method of G&S (non-binary trees, topic at the root). The two judges agreed on all but one of the segment boundaries. In what follows, "new sister" refers to adding a new sister node to the discourse structure tree as in adding node 3 to (16a), resulting in (16b).



"New subsegment" refers to adding node 4 to the tree in (16a), resulting in the tree of (17a). In terms of text, a new sister looks like the bottom portion of (17b), and a new subsegment like the innermost portion of (17c).



The author then examined factors which might cause the reader to segment the discourse at new segment boundaries, as well as factors which are present when the reader does not change segment. The cues which seem likely to be important in signalling a new discourse segment are new paragraph indentation, discourse cues such as "Turning to...", "In summary", the use of event anaphors and changes in coherence relation, sentence subject and tense. We wanted to see whether pronoun use could be predicted by segmentation algorithm which ignored pronoun use in locating segment boundaries. Table 1 shows the results.

Table 1. Signals of New Segments in a 16000 word corpus

	New Sister Segment	New Subsegment	Non-segmenting
Change in coherence relation	92%	88%	<1%
Change of sentence subject	89%	70%	52%
Segmenting cue phrase	16%	11%	—
Event anaphora	29%	9%	13%
Change in tense or aspect	49%	53%	30%

The most consistent correlate of new segment was change in coherence relation. Change in coherence relation is a stronger predictor than change in sentence subject, because while the latter is present in 89% of new sister segments and 70% of new subsegments, it is also present in half of the clauses which do not change segment. In contrast, change of coherence relation is almost never present when there is no change in segment. It was surprising that cue phrases are relatively unimportant, though clear, indicators of change of segment. Segmenting cue phrases were present in only 16% of the clauses which changed segment. On the other hand, event anaphors proved to be more important. Event anaphors are pronouns such as "this" or "it", or definite NP's such as "the move", which have events as antecedents. Frequently the antecedents are segments of the discourse, and the event anaphor makes the reader segment the previous discourse correspondingly. (This fact supports Asher's (1993) argument that event anaphoric phenomena are indicative of the processes involved in discourse segmentation.) Change in tense or aspect is not an informative factor in segmentation, except insofar as it contributes to coherence relation assignment.

Anaphoric devices (both eventive and non-eventive) are also important cues in segmentation. Our Study 2 reveals several important distinctions turning on type of anaphoric device and type of referent. Individual type referents of anaphors must be inside the segment, but event type referents tend to be outside it. Furthermore, definite NP anaphors can refer anywhere in the text, but the further away the antecedent the "heavier" the anaphoric NP. The study also confirms G&S's point that pronouns do not refer outside the focus stack (that is, they refer either inside the segment or to an entity introduced in a dominating clause in the discourse). In fact, we found that all individual pronouns in the WSJ corpus had referents in the immediately preceding sentence inside the segment. We studied

the positions of anaphoric devices and their antecedents in the 16000 words of WSJ corpus which we segmented. Our findings were:

- Pronouns had antecedents in the immediately preceding sentence in the same segment.
- Demonstratives with event antecedents had antecedents outside the segment, but demonstratives with individual antecedents did not.
- Definite NP's have antecedents in other segments.
- There is a correlation between the heaviness of an anaphor and the distance (in number of sentences) back to its antecedent.

The heaviness hierarchy is:

- pronoun
- Det NP
- Det (Adj) NP (PP)
- Det (Adj) NP (Participial)
- Det (Adj) NP (Relative Clause)

First, these findings corroborate Asher's (1993) fundamental distinction between event and individual anaphora. Here the constraints on event and individual anaphora are very different. Second, these findings support a method of segmentation to be described below, because they call for constraints on anaphors which can be simply stated with relation to the proposed tree structure. The constraints are:

Constraint on Individual Anaphora

The antecedent must be in local sister leaf node (minimal governing node which c-commands the node with the pronoun), or in the global topic.

Constraint on Abstract Type Anaphora

The antecedent must be in a sister, not necessarily local, or some dominating node.

Constraint on Heavy Definite NP Anaphora

Unconstrained.

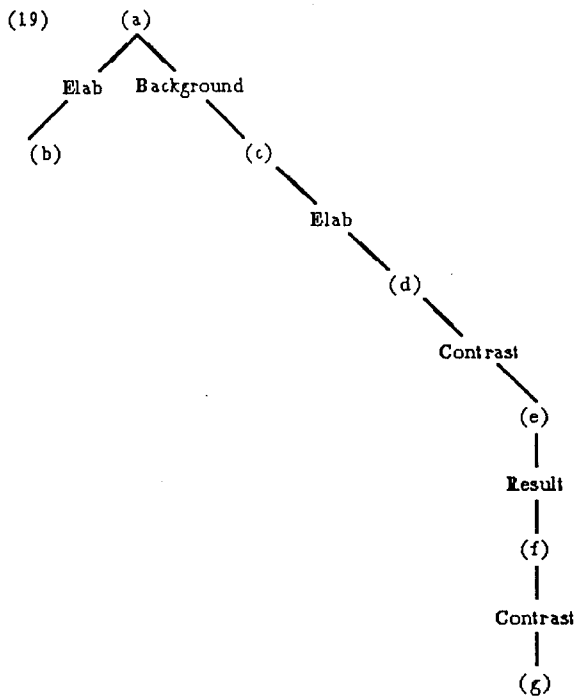
Thus, abstract type anaphors may find their antecedents up the tree at any point, while individual anaphors must find their antecedents locally or at the root. These constraints are modified below in light of the revised discourse structure theory. We also found that the notion of topic was substantiated by the facts of anaphor resolution in the WSJ corpus. Lord counted the distance back to the antecedent for all anaphors (pronouns, demonstratives, and definite NPs), using 23 topic event, 13 non-topic event, 84 topic event participant, and 80 non-topic event participants. The generalization held that antecedents in topic events and topic event participants could be located farther away than other antecedents. Looking at event anaphors, for non-topics, the antecedent was typically in the previous sentence and in the same segment. For topics, the antecedent was sometimes in the previous sentence, but more often back two or more sentences, and in a previous segment. Looking at individual anaphors, the antecedent was typically in the previous sentence and the same segment, but the likelihood of its being farther back was greater for topic event participants than for non-topic event participants.

4. Improving Our Proposal

The trees we have presented so far do not tell the whole story of discourse structure. We would like to have the trees reflect the anaphora resolution constraints. In our corpus, all individual anaphors had antecedents in the previous sentence, and all but two of the antecedents were inside the same segment. However, we know that in general anaphoric pronouns are not as constrained in their use as that. It is easy to imagine a segment in which an individual is introduced, not a topic participant, followed by a complex rhetorical structure, such as an argument, followed by a sentence with a personal pronoun. G&S introduce focus spaces, and Reichman (1985) context spaces, on the theory that topic-related (or plan-related) segments constrain anaphor resolution. In their examples, there is no case of a pronoun in a sister or closed focus space. We see no reason, in expository text (as opposed to task-oriented dialogue), that such examples should not be found, as long as a certain general distance and same-gender constraints are upheld. In task-oriented dialogue, it makes sense that certain objects which are no longer relevant because they have been set down on the work bench, as in the pump example (G&S), should not be available. However, in expository text, or narrative, it is possible to introduce a person or thing, continue on a related or background subject or argument, and then refer to the introduced person or thing with a pronoun. Consider the following text:

(18) (a) John likes to invent computer games.
(b) He often collaborates with his sister to come up with unusual logics.
(c) He started out with games like Dig Dug and Mario.
(d) These games fascinated him,
(e) but after a short time he would pick up the logical structure, and
(f) become bored.
(g) She felt the same way...

In this example, John is a topic participant, but his sister isn't. Using the trees we've been proposing so far, (18) is analyzed as in (19).

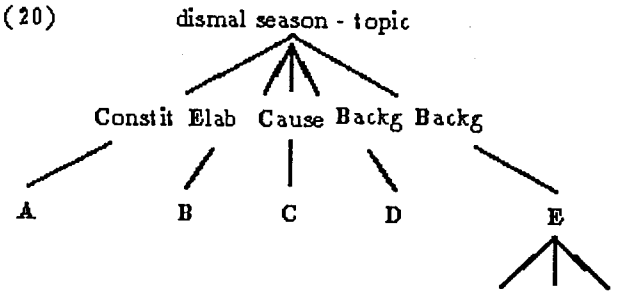


The antecedent of "She" in clause (g) is found in clause (b). Thus (18) is a counterexample to our constraint on individual pronominal anaphora resolution. In G&S terms, the segment labeled (b) in (19) should have popped from the focus stack because it does not dominate the segment starting at (c). Though (18) is a counterexample to the constraint, some other constraints must replace it. We know that the antecedent has to be close enough for human memory processing to be able to recover the antecedent. Further, if same-gender, commonsensically plausible alternatives intervene, it becomes hard for the hearer to recover the antecedent. On the other hand, it is clear that a rule which examines finely grained local-coherence based discourse trees cannot capture the empirical facts concerning pronominal anaphora resolution constraints, which allow for looking back into "popped" local segments.

An independent reason to reject clause-based or propositional content of clause-based discourse structure theory is that it ignores the distinction between semantics and discourse structure. Embedding phenomena, including belief contexts, conditionals and others, universal quantifiers, and any other phenomena with more than one clause in a sentence, must be taken into account by the discourse structure representation in a clause-based theory.

In considering how to alter our theory, we must remember that text segmentation varies with genre, and the example from task oriented dialogue discussed by G&S is unusual in having extremely localized focus spaces. In other genres, such as commentary, expository pedagogy (textbooks), and narrative fiction, segmentation operates to chunk text into topic-related segments. The typical novel segments into chapters. Chapters can have (though they need not have) topics dealing with background, action sequences, persons and so on. Within the chapter there are smaller segments, some of which may have topics. These may also have subsegments, forming a dominance hierarchy. The point is that unlike task-oriented dialogue, inside certain-sized chucks, anaphoric pronouns don't function according to the focus space constraints suggested by G&S, as exemplified in (18).

Our discourse trees must be relatively shallow, then, in order to account for availability. Both length and genre must be considered in segmentation. We need a theory which says that in the commentary genre, segments are topic-related, dominance in the discourse tree is topic dominance, and leaves of the discourse structure tree are chunks of text of a certain length (determined by memory processing constraints, and correlated with the possibility of recovering antecedents of individual pronominal anaphora). For a commentary text like that summarized in Section 4.4, this means a discourse structure like (20).



The segments labeled A–D are summed events (states, propositions, etc.) and the length in the surface structure is short enough that personal pronouns can find their antecedents within them. In the text of Section 4.4, segment E contains relatively long subsegments dealing with the decline of the three networks. Each subsegment has an elaborative structure in which pronouns can find their antecedents. (Noting, of course, that in this particular text, personal pronouns all had antecedents in the preceding sentence.) Thus, our theory must recognize length in segmentation, as well as a recursive topic related structure. Exactly the length limit for a segments which predict individual pronoun anaphora is an empirical question concerning human memory processing constraints.

4.1 Event Summation

Asher (1993) proposes a semantic treatment of abstract objects including events, states, facts, propositions and concepts. He shows that the pronouns "it", "this" and "that" can have as antecedents chunks of discourse which are summed events (or sums of other abstract objects). An example of such an anaphoric usage is found in the text in Section 4.4, where the demonstrative "so" has as antecedent an event type corresponding to "the kind of thing that happened to the Tattingers show". Such summing event anaphors abound in our corpus of commentary text. Asher also shows that events, propositions and other abstract types are antecedents of pronouns more locally, and proposes a theory which also accounts for various syntactic phenomena such as verb phrase ellipsis. He demonstrates that event summation is a necessary part of any semantic theory. He then goes on to define event summation semantically within Discourse Representation Theory (Kamp 1981), as the join of discourse events which can plausibly occur in an event chain. Built into the theory is the recognition of naive semantic constraints on event summation. A felicitous discourse will not force the hearer to sum events which in naive semantics are unlikely to be related. An example of such an infelicitous discourse is found in (21).

(21) John invested heavily. Then he ate pizza. It was gross.

The cue phrase "then" invites the summation of the investing and the pizza eating, and if so summed, the antecedent of "it" should be the sum of the two events, but since these two events are difficult to picture as connected in any way, the discourse is infelicitous.

4.2 Revised Discourse Structure Theory

Our revised discourse structure theory is as follows:

> The propositional content of discourse segments is placed at nodes. The length of a discourse segment is genre-relative. In task-oriented dialogue, discourse segments are any change of coherence relation, while in commentary and narrative, length is a size chunk in which the availability of antecedents for personal pronouns holds. Each node is labeled with a coherence relation involving the content at that node, and a higher node. Thus arcs in the tree signify that the nodes they connect are coherence related. Height in the tree corresponds to topic-relatedness in some genres. The deeper the tree, the more complex the outline, and the longer the discourse. Subtrees in topic genres are subtopics. The global topic of the whole discourse is at the root of the tree. Internal to the discourse segment in certain genres is recursively-built coherence structure among abstract types which does not constrain anaphora resolution. Interruptions and repairs are not part of the discourse structure tree.

This theory differs from the clausal version. The changes incorporate event summation, and shallowness of discourse structure to account for anaphora resolution constraints. The resulting anaphor resolution constraints reflecting the revisions in definition of the trees are given below.

Constraint on Individual Anaphora

The antecedent must be inside the same segment, or in the global topic.

Constraint on Event (Abstract Type) Anaphora

The antecedent must be in a local sister or in a dominating node.

4.3 Formal Discourse Structure

Each clause introduces individual, event, state, and propositional reference markers into the discourse representation structure (DRS). (22) is the DRS representing the text "John invested heavily. He made a huge profit". Temporal relations between events are reflected in the DRS, and anaphoric relations are determined. Coherence can be viewed as defined over constituents of the DRS which reflect the content introduced by clauses.

(22)

u1,e1,u2,u3,e2,r1,r2
r1 < now
john(u1)
e1 invest(u1)
e1 incl in r1
heavily(e1)
profit(u3)
e2 make(u2,u3)
huge(u3)
r1 < r2
e2 included in r2
u2=u1

The constituent K1 corresponding to John's investing is as in (23). A Discourse Structure (DS) represents the coherence of a DRS.

(23)

u1,e1,r1
r1 < now
john(u1)
e1 invest(u1)
e1 incl in r1
heavily(e1)

In (24) we see the DS for (23). Here the two constituent DRS's introduced by John's investing (K1) and John's making a profit (K2) are causally related.

(24)

topic(K1,goal(K1,K2))

In order to integrate a DRT approach with the principles stated in Section 4.2, we must again revise those principles:

Definition of Discourse Segment

A node of the discourse structure is a portion of a DRS, that is, some (possibly non-contiguous in surface structure) subset of the reference markers and corresponding conditions in the DRS.

The content of that constituent DRS plays a topic role in the discourse. The constituent DRS comes from a portion of the surface discourse which does not exceed maximum memory capacity for finding antecedents of personal pronouns.

Arcs are labeled with binary coherence relations.

The following principle must be added:

Discourse segment boundaries occur where there is a change in coherence relation at the segment level.

That is, where it is no longer possible to continue summing the events, states and other abstract reference markers in such a way that the sum has the same coherence relation as before to the remainder of the discourse (in particular, to the node above in the discourse structure tree), there is a discourse segment boundary. Reflecting these changes, the anaphora resolution constraints become:

Constraint on Individual Pronominal Anaphora:

Antecedent reference markers must be in an accessible position inside the same segment, or in the global topic.

Constraint on Abstract Type Pronominal Anaphora

Antecedents may be any abstract type reference marker (event, state, fact, proposition), concept type reference markers (delineated DRS's), DRS's or discourse segments found in a local sister, or in a dominating node in the discourse structure tree.

This theory predicts that change of discourse relation results in change of segment, one of the main findings in our Study 1.

The approach proposed here predicts the availability of antecedents for anaphors by both the formal semantic constraints of Discourse Representation Theory, and the constraints of the discourse segment structure. Segments are signalled by coherence relations and coherence is defined in terms of naive

theories of event relatedness. The definition of local coherence is essentially psychological (cf. Graesser and Clark 1985b).

Global coherence is defined by essentially the same intuitions. People do segment discourse, especially certain genres, they do infer causal relations between the events reported in chunks of discourse. Furthermore, written expository text is planned and outlined by writers with segments in mind (and often explicitly marked with subtitles).

The coherence algorithm draws upon local linguistic cues plus the naive semantic lexicon. The segmentation algorithm draws upon local coherence information. Change of segment hypotheses are generated when one of the significant factors is encountered. These factors, in order of importance are: change of local coherence, change of subject, paragraph indentation, segmenting event anaphoric device segmenting cue phrase.

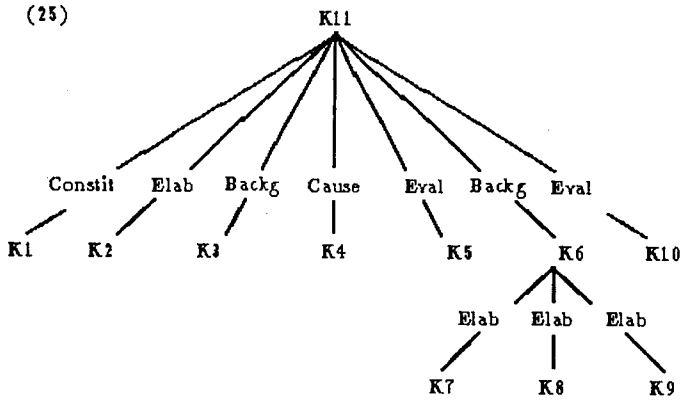
4.4 An Example

An example analysis of a text according to the final theory will be given here regarding the text sketched in the following table. (Space does not permit reproduction of the entire text). The topic-related segments are identified by discourse variables K1–K10. These variables correspond to subsets of the reference markers and conditions in the root DRS for the entire text. K1 would identify the reference markers and conditions for the content of sentences 1–5, and so on as listed below. K11 corresponds to the content of the topic sentence, sentence 6.

Vars	Sents	Text topics
K1	1–5	The promising Tattinger's show was canceled.
K2	6–10	So it goes in this dismal network season.
K3	11–14	This is taking place while television watching is on the rise.
K4	15–22	Fall debuts were delayed by the writer's strike.
K5	23–30	The studios may save millions from concessions by the strikers.
K6	31–34	This is the latest in a string of declines in network ratings.
K7	32...	CBS is hurting the worst.
K8	33...	No. 1 NBC has seen its audience shrink 7% overall.
K9	34...	ABC is the only network doing better.
K10	35–41	The strike's impact will be felt well into the next TV season.
K11	6	So it goes in this dismal network season.

The corresponding discourse structure tree would be as in (25).

(25)



5. Conclusion

We hope to have demonstrated that coherence can be explained in terms of cognitive strategies which are used to make sense of events and event chains, and structures of other types such as states, mental states, propositions and so on. A plausible architecture is one in which coherence inferences operate upon the logical form of a discourse and employ commonsense knowledge. Local coherence relations can be extracted using syntactic, semantic and naive semantic information. Knowledge representation for naive semantics (Dahlgren 1988a, 1991, 1994) is adequate to drive almost all of the coherence relation assignments, including implicit ones (Dahlgren 1981). Global discourse structure is genre relative (Polanyi 1988) and topic-related (Van Dijk and Kintsch 1983). Constraints on anaphora resolution recognize topic-related discourse segments. Hence global discourse structure trees must be relatively shallow.

Acknowledgments

I want to thank Nicholas Asher, Clair Chyi, Carol Lord, Art Graesser, Joyce P. McDowell, Edward P. Stabler, Jr. and Hajime Wada for discussions of these issues and commentary on this work. Lord (at Intelligent Text Processing, Inc.) and Stabler developed the parser and McDowell and Wada the DRT component of the natural language system (InQuizit) in which the coherence algorithms described here operate.

References

Asher, N. 1993. Reference to Abstract Objects in Discourse. Norwell, MA: Kluwer.

Asher, N. and Kamp, H. 1995. Discourse Structure in the Service of Lexical Disambiguation. Ms in prep.

Cohen, R. 1984. A Computational Theory of the Function of Clue Words in Argument Understanding. *Proceedings of the COLING-84 Conference*, 251–258.

Dahlgren, K. 1985a. The Cognitive Structure of Social Categories. *Cognitive Science* 9, 379–398.

Dahlgren, K. 1988a. *Naive Semantics for Natural Language Understanding*. Boston, MA: Kluwer.

Dahlgren, K. 1988b. Using Commonsense Knowledge to Disambiguate Word Senses. In P. St Dizier and V. Dahl (eds.), *Natural Language Understanding and Logic Programming 2*. Amsterdam: North-Holland.

Dahlgren, K. 1989. Coherence Relations and Naive Semantics. Paper given at the Symposium on Modelling Discourse Structure: Discourse Segments and Discourse Relations, University of Texas, Austin.

Dahlgren, K. 1990. Naive Semantics and Robust Natural Language Processing. *Proceedings of the AAAI Spring Symposium.*

Dahlgren, K. 1991. The Autonomy of Shallow Lexical Knowledge. In J. Pustejovsky (ed), *Knowledge Representation and Lexical Semantics*. New York: Springer-Verlag.

Dahlgren, K. 1992. Convergent Evidence for a Set of Coherence Relations. In D. Stein (ed.), *Cooperating with Written Texts*. New York: Mouton de Gruyter.

Dahlgren, K. 1993. A Linguistic Ontology. *Proceedings of the International Workshop on Formal Ontology.*

Dahlgren, K. 1994. Finding Relevant Texts. *Proceedings of the AAAI Fall Symposium*, New Orleans.

Dahlgren, K. and J. McDowell. 1986a. Kind Types in Knowledge Representation. *Proceedings of the COLING-8 Conference.*

Dahlgren, K., J.P. McDowell, and E.P. Stabler, Jr. 1989. Knowledge Representation for Commonsense Reasoning with Text. Forthcoming in *Computational Linguistics*.

Davidson, D. 1967a. Causal Relations. *Journal of Philosophy* 64, 692–703.

Davidson, D. 1967b. The Logical Form of Action Sentences, in N. Rescher (ed.), *The Logic of Action and Preference.*

Decker, N. 1985. The Use of Syntactic Clues in Discourse Processing. *Proceedings of the ACL-85 Conference.*

Foss, D.J. 1988. Experimental Psycholinguistics. *Annual Review of Psychology* 30, 301–348.

Graesser, A. and L. Clark. 1985a. *Structure and Procedures of Implicit Knowledge.* Norwood, NJ: Ablex.

Graesser, A. and L. Clark. 1985b. The Generation of Knowledge-Based Inferences during Narrative Comprehension. In G. Rickheit and H. Strohner (eds.), *Inferences in Text Processing*. Amsterdam: North-Holland.

Grosz, B. and C. Sidner. 1986. Attention, Intensions and the Structure of Discourse: A Review. *Computational Linguistics* 7, 85–98; 12, 175–204.

Hayes, P. J. 1985. The Second Naive Physics Manifesto. In J.R. Hobbs and R.C. Moore (eds.), *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex.

Hinds, J. 1977. Paragraph Structure and Pronominalization. *Papers in Linguistics* 10(1–2), 77–99.

Hirst, G. 1981. Discourse-Oriented Anaphora Resolution: A Review. *Computational Linguistics* 7, 85–98.
Hobbs, J.R. 1979. Why is Discourse Coherent? SRI Technical Note #176.
Hobbs, J.R. 1985. On the Coherence and Structure of Discourse. CSLI Report #CSLI-85-37.
Holland D. and N. Quinn, eds. 1987. *Cultural Models in Language and Thought.* Cambridge: Cambridge University Press.
Hovy, E.H. 1990. Parsimonious and Profligate Approaches to the Question of Discourse Structure Relations. *Proceedings of the 5th International Workshop on Text Generation.*
Johnson-Laird, P.N. 1983. *Mental Models.* Cambridge: Harvard University Press.
Kamp, H. 1981. A Theory of Truth and Semantic Representation. In J. Groenendijk, Th. Janssen, and M. Stokhof (eds.), *Formal Methods in the Study of Language.* Amsterdam: Mathematisch Centrum, 277–322.
Lakoff, G. 1985. *Women, Fire and Dangerous Things.* Chicago, IL: University of Chicago Press.
Mann, W.C. and S.A. Thompson. 1987. Rhetorical Structure Theory: A Theory of Text Organization. ISI Reprint Series: ISI-RS-87-190.
Morrow, D.G, S.L. Greenspan, and G.H. Bower. 1987. Accessibility and Situation Models in Narrative Comprehension. *Journal of Memory and Language* 26, 165–187.
Partee, B. 1984. Nominal and Temporal Anaphora. *Linguistics and Philosophy* 7, 243–286.
Polanyi, L. 1988. A Formal Model of the Structure of Discourse. *Journal of Pragmatics* 12, 601–638.
Reichman-Adar, R. 1984 Technical Discourse: The Present Progressive Tense, The Deictic 'That' and Pronominalization. *Discourse Processes* 7, 337–369.
Reiter, R. 1980. A Logic for Default Reasoning. *Artificial Intelligence* 13, 81–132.
Rosch, E., C.B. Mervis, W.D. Gray, D.M. Johnson, and P. Boyes-Braem. 1976. Basic Objects in Natural Categories. *Cognitive Psychology* 8, 382–439.
Schubert, L. K., R.G. Goebel, and N.J. Cercone. 1979. The Structure and Organization of a Semantic Net for Comprehension and Inference. In N.V. Findler (ed.), *Associative Networks.* New York: Academic Press.
Smith, E.E. and D.L. Medin. 1981. *Categories and Concepts.* Cambridge, MA: Harvard University Press.
Strawson, P.C. 1953. *Individuals.* London: Methuen.
Trabasso, T. and L.L. Sperry. 1985. Causal relatedness and the Importance of Story Events. *Journal of Memory and Language* 24, 595–611.
Trabasso, T. , T. Secco, and P. van den Broek. 1983. Causal cohesion and story coherence. In H. Mandl, N.S. Stein and T. Trabasso (eds.), *Learning and Comprehension of Text.* Hillsdale, NJ: Erlbaum.
Van Dijk, T. and W. Kintsch. 1983. *Strategies of Discourse Comprehension.* New York: Academic Press.
Vendler, Z. 1967. *Linguistics in Philosophy.* Ithaca: Cornell University Press.
Wada, H. 1994. A Treatment of Functional Definite Descriptions. *Proceedings of the Coling-94 Conference.*
Wilks, Y. 1975. Preference Semantics. In E. Keenan (ed.), *Formal Semantics of Natural Language.* Cambridge: Cambridge University Press.

Chapter 6
On the Relation Between the Informational and Intentional Perspectives on Discourse

Jerry R. Hobbs

SRI International, Menlo Park

1. Introduction

In the paper "Interpretation as Abduction" (hereafter IA) Hobbs et al. (1993 presented and elaborated the view that to interpret an utterance is to find the bes explanation of why it would be true. We may call this the "Informationa Perspective" on discourse interpretation. The only thing to be explained is th information explicitly conveyed by the utterance, and the explanation does no necessarily involve any knowledge of the specific goals of the speaker.

Norvig and Wilensky (1990) raised the objection to this approach that wha really needs to be explained is what the speaker was trying to accomplish with th utterance. Under this view, to interpret an utterance is to find the best explanatio of why it was said. We may call this the "Intentional Perspective" on discours interpretation.

The Intentional Perspective has been the canonical view in natural languag processing since the middle 1970s. It originated with Power (1974), Bruce (1975). and Schmidt et al. (1978), and is the view adopted in Cohen and Perrault (1979). Grosz (1979), Allen and Perrault (1980), Perrault and Allen (1980), Hobbs and Evans (1980), Grosz and Sidner (1986) and many others since that time. The view taken in all of this work is that the speaker is executing a plan, the utterance is an action in that plan, and the job of the hearer is to discover the plan and the role that the utterance plays in the plan. This is an especially useful, indeed essential, perspective when the discourse is a dialogue in which most turns are a sentence or less in length and the participants' plans are being modified continuously by the interaction.

It is clear why the Intentional Perspective is the correct one when we look at things from the broadest possible point of view. An intelligent agent is embedded in the world and must, at each instant, understand the current situation. The agent does so by finding an explanation for what is perceived. Put differently, the agent must explain why the complete set of observables encountered constitutes a coherent situation. Other agents in the environment are viewed as intentional, that is, as planning mechanisms, and this means that the best explanation of their observable actions is most likely to be that they are steps in a coherent plan. Thus, making sense of an environment that includes other agents entails making sense of