Handbook of Discourse Processes

Edited by

Arthur C. Graesser
University of Memphis

Morton Ann Gernsbacher
University of Wisconsin, Madison

Susan R. Goldman
University of Illinois, Chicago

LEA
2003 LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey London

12

Discourse in Computational Linguistics and Artificial Intelligence

Johanna D. Moore
Peter Wiemer-Hastings
University of Edinburgh

Models of discourse structure and processing are crucial for constructing computational systems capable of interpreting and generating natural language. Research on discourse focuses on two fundamental questions within computational linguistics and artificial intelligence. First, what information is contained in extended sequences of utterances that goes beyond the meaning of the individual utterances? Second, how does the context in which an utterance is used affect the meaning of the individual utterances or parts of them?

Discourse research in computational linguistics and artificial intelligence (AI) encompasses work on spoken and written discourse, monologues as well as dialogues (both spoken and keyboarded). The questions that discourse research attempts to answer are relevant to all combinations of these features. The juxtaposition of individual clauses may imply more than the meaning of the clauses themselves, regardless of whether the clauses were contributed by the same speaker (writer).[1] Likewise, the context created by prior utterances affects the current one regardless of which participant uttered it.

In this chapter, we first provide an overview of types of discourse structure and illustrate these with examples. We then describe the in-

[1]Henceforth, we use *speaker* and *hearer* to indicate the producer and interpreter of discourse, respectively, whether it is spoken or written. Where the distinction between spoken and written discourse is important, we are more explicit.

fluential theories that account for one or more types of structure. We then show how these theories are used to address specific discourse processing phenomena in computational systems. Finally, we discuss the use of discourse processing techniques in a range of modern language technology applications.

OVERVIEW OF DISCOURSE STRUCTURE

Researchers in computational linguistics have long argued that coherent discourse has structure, and that recognizing the structure is a crucial component of comprehending the discourse (Grosz & Sidner, 1986; Hobbs, 1993; Moore & Pollack, 1992; Stone & Webber, 1998). Interpreting referring expressions (e.g., pronouns and definite descriptions), identifying the temporal order of events (e.g., the default relationship between the falling and pushing events in "Max fell. John pushed him."), and recognizing the plans and goals of our interlocutors all require knowledge of discourse structure (Grosz & Sidner, 1986; Kehler, 1994b; Lascarides & Asher, 1993; Litman & Allen, 1987). Moreover, early research in language generation showed that producing natural-sounding multisentential texts required the ability to select and organize content according to rules governing discourse structure and coherence (Hovy, 1988b; McKeown, 1985; Moore & Paris, 1993).

Although there is still considerable debate about the exact nature of discourse structure and how it is recognized, there is a growing consensus among researchers in computational linguistics that at least three types of structure are needed in computational models of discourse processing (Grosz & Sidner, 1986; Hobbs, 1993; Moore & Pollack, 1992). These are described next.

Intentional structure describes the roles that utterances play in the speaker's communicative plan to achieve desired effects on the hearer's mental state or the conversational record (Lewis, 1979; Thomason, 1990). Intentions encode what the speaker was trying to accomplish with a given portion of discourse. Many have argued that the coherence of discourse derives from the intentions of speakers, and that understanding depends on recognition of those intentions (e.g., Grice, 1957; Grosz & Sidner, 1986). Research in response generation shows that, to participate in a dialogue, agents must have a representation of the intentional structure of the utterances they produce. Intentional structure is crucial for responding effectively to questions that address a previous utterance: without a record of what an utterance was intended to achieve, it is impossible to elaborate or clarify that utterance (Moore, 1995; Young, Moore, & Pollack, 1994a). Moreover, speaker intentions are an important factor in generating nominal expressions (Appelt, 1985; Green, Carenini, & Moore, 1998) and selecting appropriate lexical items, including discourse cues (e.g., *because, thus*; Moser & Moore, 1995; Webber, Knott, Stone, & Joshi, 1999) and scalar terms (e.g., *difficult, easy*; Elhadad, 1995).

Informational structure consists of the semantic relationships between the information conveyed by successive utterances (Moore & Pollack, 1992). Causal relations are a typical example of informational structure, and psychologists working in reading comprehension have shown that these relations are inferred during reading (Gernsbacher, 1990; Graesser, Singer, & Trabasso, 1994; Singer, Revlin, & Halldorson, 1992). In addition, several researchers identified types of text whose organization follows the inherent structure of the subject matter being communicated (e.g., the structure of the domain plan being discussed; Grosz, 1974; Linde & Goguen, 1978) or the spatial (Sibun, 1992; Linde, 1974), familial (Sibun, 1992) or causal relationships (Paris, 1988; Suthers, 1991) between the objects or events being described, or the states and events being narrated (Lehnert, 1981). Several systems that generate coherent texts based on domain or informational structure have been constructed (Paris, 1988; Sibun, 1992; Suthers, 1991).

Attentional structure as defined by Grosz and Sidner (1986) contains information about the objects, properties, relations, and discourse intentions that are most salient at any given point in the discourse. In discourse, humans focus or center their attention on a small set of entities and attention shifts to new entities in predictable ways. Natural language understanding systems must track attentional shifts to resolve anaphoric expressions (Gordon, Grosz, & Gilliom, 1993; Grosz, 1977; Sidner, 1979) and understand ellipsis (Carberry, 1983; Kehler, 1994a). Natural language generation systems track focus of attention as the discourse as a whole progresses as well as during the construction of individual responses to influence choices on what to say next (Kibble, 1999; McCoy & Cheng, 1990; McKeown, 1985), determine when to pronominalize (Elhadad, 1992), make choices in syntactic form (e.g., active vs. passive; Elhadad, 1992; McKeown, 1985; Mittal, Moore, Carenini, & Roth, 1998), appropriately mark changes in topic (Cawsey, 1993), and generate elliptical utterances.

In addition to these three primary types of discourse structure, the literature on discourse in computational linguistics has discussed two additional types of structure. One of them, rhetorical structure, has had considerable impact on computational work in natural language generation.

Information structure consists of two dimensions: (a) the contrast a speaker makes between the part of an utterance that connects it to the rest of the discourse (the *theme*), and the part of an utterance that contributes new information on that theme (the *rheme*); and (b) what the speaker takes to be in contrast with things a hearer is or can be attending to. Information structure can be conveyed by syntactic, prosodic, or morphological means. Steedman argued that information structure is the component of linguistic structure (or grammar) that links intentional and attentional structure to syntax and prosody via compositional semantics for notions like theme (or topic) and rheme (or comment). Recently, a number of theories of information structure (Steedman, 1991; Vallduvi, 1990) have brought hitherto unformalized notions like theme, rheme, and focus within the compositional semantics that forms a part of formal grammar (Steedman, 2000, 2001).

Rhetorical structure is used by many researchers in computational linguistics to explain a wide range of discourse phenomena. There have been several proposals defining the set of *rhetorical* (or *discourse* or *coherence*) relations that can hold between adjacent discourse elements and researchers have attempted to explain the inferences that arise when a particular relation holds between two discourse entities, even if that relation is not explicitly signaled in the text. Researchers in interpretation have argued that recognizing these relationships is crucial for explaining discourse coherence, resolving anaphora, and computing conversational implicature (Hobbs, 1983; Lascarides & Asher, 1993; Mann & Thompson, 1988). Researchers in generation have shown that it is crucial for a system to recognize the additional inferences that are conveyed by the sequence of clauses they generate because these additional inferences may be the source of problems if the user does not understand or accept the system's utterance. Moreover, to implement generation systems capable of synthesizing coherent multisentential texts, researchers identified patterns of such relations that characterize the structure of texts that achieve given discourse purposes, and many text generation systems have used these patterns to construct coherent monologic texts to achieve a variety of discourse purposes (Hovy, 1991; McKeown, 1985; Mellish, O'Donnell, Oberlander, & Knott, 1998; Mittal et al., 1998; Moore & Paris, 1993; Rösner & Stede, 1992; Scott & de Souza, 1990).[2]

Much of the remaining debate concerning discourse structure within computational linguistics centers around which of these structures are primary and which are parasitic, what role the structures play in different discourse interpretation and generation tasks, and whether their importance or function varies with the discourse genre under consideration. For example, a major result of early work in discourse was the determination that discourses divide into segments much like sentences divide into phrases. Each utterance of a discourse either contributes to the preceding utterances or initiates a new unit of meaning that subsequent utterances may augment. The usage of a wide range of lexicogrammatical devices correlates with discourse structure, and the meaning of a segment encompasses more than the meaning of the individual parts. In addition, recent studies show significant agreement among segmentations performed by naive subjects (Passonneau & Litman, 1997). Discourse theories differ about the factors they consider central to explaining this segmentation and the way in which utterances in a segment convey more than the sum of the parts. Grosz and Sidner (1986) argued that intentions are the primary determiners of discourse segmentation, and that linguistic structure (i.e., segment embedding) and attentional structure (i.e., global focus) are dictated by relations between intentions. Polanyi (1988) took an opposing view and claimed that hierarchical structure "emerges from the structural and semantic relationships obtaining among the linguistic units which speakers use to build up their discourses" (p. 602). In Hobbs (1985), segmental structure is an artifact of binary *coherence relations* (e.g., background, explanation, elaboration) between a current utterance and the preceding discourse. Despite these different views, there is general agreement concerning the implications of segmentation for language processing. For example, segment boundaries must be detected to resolve anaphoric expressions (Asher, 1993; Grosz & Sidner, 1986; Hobbs, 1979; Passonneau & Litman, 1997). Moreover, several studies have found prosodic as well as textual correlations with segment boundaries in spoken language (Grosz & Hirschberg, 1992; Hirschberg, Nakatini, & Grosz, 1995; Nakatani, 1997; Ostendorf & Swerts, 1995), and appropriate usage of these intonational indicators can be used to improve the quality of speech synthesis (Davis & Hirschberg, 1988).

In the sections that follow, we further examine the theories of discourse structure and processing that have had significant impact on computational models of discourse phenomena. A comprehensive survey of discourse structure for natural language understanding appears in Grosz, Pollack, and Sidner (1989), and therefore we focus on discourse generation and dialogue in this chapter. We then review the role that discourse structure and its processing play in a variety of current natural language applications. The survey in Grosz et al. (1989)

[2]Moore and Pollack (1992) argued that the rhetorical relations used in these systems typically conflate informational and intentional considerations, and thus do not represent a fourth type of structure.

focuses largely on the knowledge-intensive techniques prevalent in the late 1980s. Here we emphasize statistical and shallow processing approaches that enable discourse information to be used in a wide range of today's natural language technologies.

COMPUTATIONAL THEORIES OF DISCOURSE STRUCTURE AND SEMANTICS

Discourse Representation Theory

Discourse Representation Theory (DRT) is a formal semantic model of the processing of text in context that has applications in discourse understanding. DRT was originally formulated in Kamp (1981) and further developed in Kamp and Reyle (1993), with a concise technical summary in van Eijck and Kamp (1997). DRT grew out of Montague's model-theoretic semantics (Thomason, 1974), which represents the meanings of utterances as logical forms and supports the calculation of the truth conditions of an utterance. DRT addresses a number of difficulties in text understanding (e.g., anaphora resolution) that act at the level of the discourse.

This section gives a brief overview of the philosophy behind DRT, the types of structures and rules that DRT uses, and the particular problems that it addresses. We also describe some of the limitations of the standard DRT theory.

Philosophical Foundations of DRT. As mentioned earlier, DRT is concerned with ascertaining the semantic truth conditions of a discourse. The semantic aspects of a discourse are related to the meaning of the discourse, but not related to the particular situation (including time, location, common ground, etc.) in which the discourse is uttered. The advantage to this approach from the logical point of view is that the semantic representation for the discourse can be automatically (more or less) built up from the contents (words) and structure of the discourse alone without bringing in information about the external context of the utterance. Once constructed, it can be compared with a logical representation of some world (a *model* in DRT terms) to determine whether the discourse is true with respect to that model.

DRT Structures. The standard representation format in DRT, known as a discourse representation structure (DRS), consists of a box with two parts as shown in Fig. 12.1. The top part of the box lists the discourse referents, which act as variables that can be bound to different

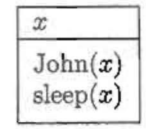


FIG. 12.1. A simple DRS for the sentence, "John sleeps."

entities in the world. The bottom section of the DRS lists the propositions claimed to be true of those referents in the described situation. Figure 12.1 gives the DRS of the sentence "John sleeps." The representation can be read as, "There is an individual who is named *John*, and of whom the *sleep* predicate is true." This is equivalent to the logical expression: $(x) : John(x) \quad sleep(x)$.

DRT Rules and Processing. To derive a structure like the one shown here, DRT uses a set of standard context-free grammar rules and a set of semantic interpretation rules based on the syntactic structure of the input sentence. Figure 12.2 shows a simple DRT rule for processing proper nouns. The left-hand side shows a segment of the syntactic tree that must be matched, and the right-hand side shows the result of applying the rule, including adding propositions to the DRS and changing the parse tree. This rule applies to the structure on the left in Fig. 12.3, which shows the parse tree of the example sentence, "John sleeps," in a DRS. The rule produces the representation on the right in Fig. 12.3 by deleting part of the parse tree, inserting a variable in its place, and adding a proposition to the DRS.

Next, a similar rule is applied that reduces the verb phrase to a proposition, *sleep*, which is true of a new discourse referent, y. Then the sentence rule deletes the remaining syntactic structure and equates the subject discourse referent with the object referent, $x = y$. Finally, x is substituted for y in the *sleep* proposition, producing the structure shown in Fig. 12.1. The next sentence in a discourse is proc-

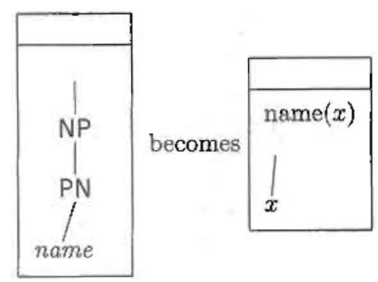


FIG. 12.2. A DRS rule processing proper nouns.

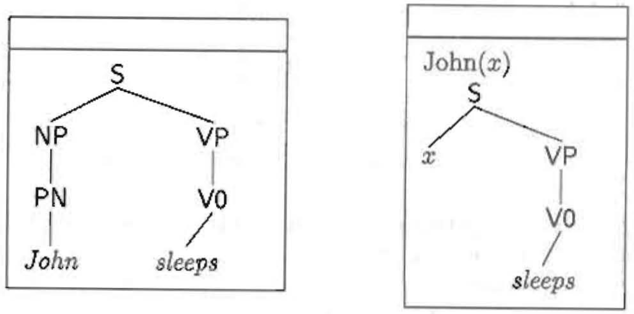


FIG. 12.3. Before and after applying the proper noun rule.

essed by adding its parse tree to the box and applying the semantic transformation rules to it.

The construction of a complete DRS enables the calculation of its truth value with respect to a model. A model is a formal representation of the state of the world. Figure 12.4 shows a model distinguished from a DRS by its double box. The elements in the top section of a model are interpreted differently than those in a DRS. In a DRS, they are variables that can bind to entities in the world. In a model, each referent indexes a particular entity in the world.

In this example, the DRS is true with respect to the model because there is a consistent mapping (x = b) between the discourse referents in the DRS and the individuals in the model, and the model contains all of the propositions that are in the DRS. Because the model is taken to be a snapshot of the world, it may contain many additional propositions not in the DRS without affecting the truth conditions of the DRS. Those that are not relevant to the DRS are simply ignored.

Uses of DRT. The major advantages of DRT are that it provides a simple, structure-based procedure for converting a syntactic representation of a sentence into a semantic one, and that semantic repre-

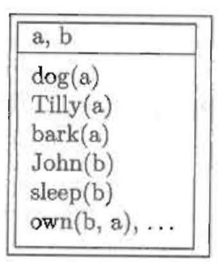


FIG. 12.4. A snapshot of the world.

sentation can be mechanically compared to a representation of the world to compute the truth-conditional status of the text. DRT also addresses (at least partially) the difficult discourse problems of anaphora resolution, quantifier scoping, and presupposition.

When a pronoun is processed in DRT, the semantic interpretation rule adds an instruction of the form x = ?, which is read as, "find some discourse referent x in the discourse context." The referent must satisfy three constraints: consistency, structural, and knowledge. The consistency constraint specifies that the new mapping of a discourse referent must not introduce a contradiction into the DRS. In practice, this ensures that number and gender restrictions are applied. The structural constraint limits where in a complex DRS a coreferent can be found. In practice, this is similar to the constraints proposed in centering theory (Grosz, Joshi, & Weinstein, 1995). Finally, the knowledge constraint is intended to prohibit the inclusion of any coreference that would violate world knowledge or common sense. Unfortunately, the scope of this constraint makes a complete implementation of it impossible.

DRT is the foundation of recent research in psycholinguistics, which attempts to model human judgments of the acceptability of a range of anaphors. Gordon and Hendrick (1997) collected ratings from humans of the acceptability of various combinations of names, quantified, definite and indefinite noun phrases, and pronouns. Their results show that human judgments did not support some of the constraints on coreference acceptability that came from classical binding theory (Chomsky, 1981). Gordon and Hendrick (1998) claimed that a model of coreference based on DRT corresponds better with human acceptability judgments.

Of course pronominal coreference is not the only type of anaphora. Asher (1993) addressed other types of anaphora as described later. In sentences like (1), there is a structural ambiguity concerning the scope of the quantifier *every*. Specifically, there are two readings of the sentence: one in which each farmer owns a different donkey, and one in which all the farmers collectively own a particular donkey. When processing such a sentence with DRT, there is a choice of processing the quantifier before or after processing the indefinite noun phrase. The two orders of rule application produce the two different structures shown in Fig. 12.5.

(1) Every farmer owns a donkey.

Both of these DRSs include substructures that represent the quantifier as a conditional. They are read as, "if the conditions on the left

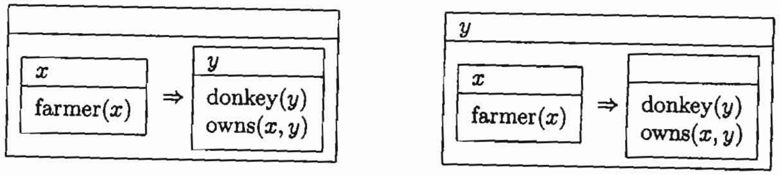


FIG. 12.5. Two readings of an ambiguous quantifier scoping.

hold, the conditions on the right must also hold." In the DRS on the left, the donkey is within the scope of the conditional. Thus, for every farmer, there should be a (potentially different) donkey. In the DRS on the right, the referent for the donkey is global and outside the scope of the conditional. Thus, there is one donkey that every farmer owns. Although this example applies only within a sentence, it suggests how hierarchical discourse relations can be represented by variants of DRT as described next.

DRT's treatment of presupposition in discourse is related to the way it handles quantifier scoping. In particular utterances such as (2a), certain propositions are said to *project* out of the sentence—that is, they are held to be true regardless of whether the premise of the conditional is. In (2a), there is a presupposition (at least for rhetorical purposes) that John has a dog, and that presupposition is true regardless of whether she has fleas. For constructs like "John's dog," DRT creates discourse referents at the global level of the representation that correspond to John and his dog.

(2) a. If John's dog has fleas, she scratches them.
b. She wears a flea collar though.
b . They jump around a lot.

The DRS for (2a) is shown in Fig. 12.6 and is similar in structure to the one shown on the left in Fig. 12.5. Here the discourse referent for

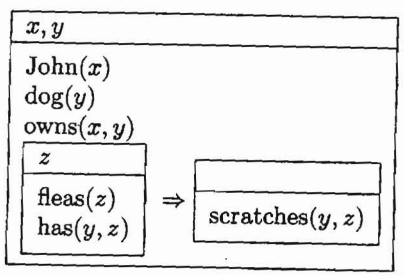


FIG. 12.6. A DRS for a conditional sentence with accommodated presuppositions.

the fleas is located within the internal left-hand box for the conditional structure. Thus, a follow-up sentence like (2b) provides no difficulty for anaphora resolution. But a sentence like (2b) is not acceptable; the referent for the fleas is inaccessible because it is embedded within a constituent DRS (Asher & Lascarides, 1998).

Limitations. Although it addresses the issues described earlier, DRT is somewhat limited as a theory of *discourse*. Most of its mechanisms address within-sentence processing. The addition of new sentences to a DRS is done in a simple way. The syntactic structure for the new sentence is added to the current DRS, and the semantic construction rules convert it to additional conditions. There is no accounting for the relationships that apply between utterances.

One direction of current research is on the tense and aspect of verbs in sentences (Kamp & Reyle, 1993). This involves the addition of a discourse referent that refers to an event. A sentence like "John slept" is represented with the predicate *sleep*($t1, j$), with the additional information $t1$ n, where n refers to the current time (now). Thus, this variant of DRT allows the explicit representation of the temporal relationships between sentences in a discourse.

Another difficulty for standard DRT is anaphora with plural referents. There is no simple mapping between pronouns and plural referents. Instead, additional inference is required to produce appropriate referents. In (3), the pronoun *they* refers not to any particular student, but to the entire set of students. To address this problem, Kamp and Reyle (1993) extended DRT to deal with conjunctions and quantifiers like *most*.

(3) Every student passed the exam. They were very happy.

In discourses like (4), the pronoun *it* does not refer to an entity, but to the fact that John failed the exam. Asher (1993) took this as a starting point for his research, which deals with reference to abstract objects. The current formalization of this research is Segmented DRT, which includes rhetorical relations between sentences; it is described later.

(4) John failed the exam, and it didn't make him happy.

Grosz and Sidner Theory

In Grosz and Sidner's (1986) theory of discourse (henceforth G&S), discourse structure is a composite of three interrelated structures called *intentional*, *attentional*, and *linguistic structure*. In G&S, in-

tentional structure consists of *discourse segment purposes* and the relationships between them. Discourse segment purposes reflect the communicative intentions of the discourse participants and lead to the utterances produced. Discourse segment purposes are thus an extension of the intentions in Grice's (1957) theory of utterance meaning: They achieve their desired effect in part by being recognized.

In G&S, a discourse segment is a group of utterances that function together to realize a speaker's purpose. A segment DS_n originates with the speaker's intention: It is exactly those utterances that the speaker produces to satisfy a communicative intention I_n in the intentional structure. In other words, I_n is the discourse segment purpose (DSP) of DS_n. The discourse structure is a hierarchy of segments originating with the structure of the speaker's intentions when producing the discourse. G&S defines two relations that may hold among DSPs. A purpose I_m *dominates* another purpose I_n when fulfilling I_n is part of fulfilling I_m. A purpose I_n *satisfaction-precedes* another purpose I_m when I_n must be satisfied first. The dominance and satisfaction-precedence relations impose a hierarchical structure on the speaker's intentions, and this in turn determines the linguistic structure of discourse. DS_n is embedded in another segment DS_m just when the purposes of the two segments are in the dominance relation (i.e., I_m dominates I_n). The dominance relation among intentions fully determines the embeddedness relations of the discourse segments that realize them.

As an example, consider the discourse shown in Fig. 12.7 adapted from Mann and Thompson (1988). The whole discourse is a segment, DS_0, that attempts to realize I_0, the speaker's intention for the hearer to adopt the intention of attending the party. As part of her plan to achieve I_0, the speaker generates I_1, the intention for the hearer to adopt the belief that there will be lots of good food. Then as part of her plan to achieve I_1, the speaker generates I_2, the intention that the hearer be-

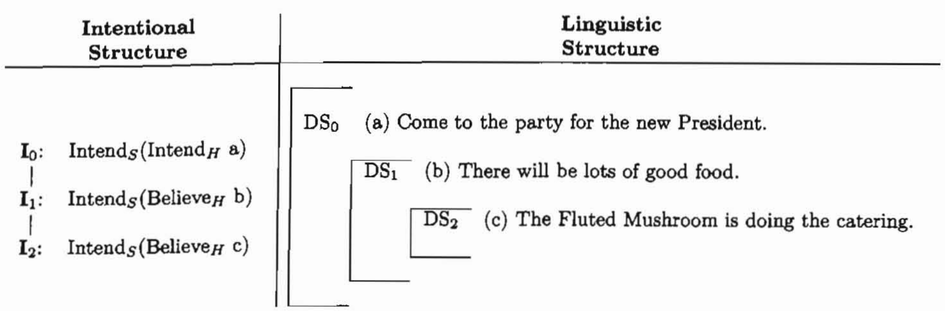


FIG. 12.7. For G&S, dominance in intentional structure determines embedding in linguistic structure.

lieve that a mutually known good caterer is providing the food. As shown on the left in Fig. 12.7, I_0 dominates I_1, which in turn dominates I_2. Due to these dominance relations, the discourse segment that realizes I_2 is embedded in the discourse segment for I_1, which is in turn embedded within the discourse segment for I_0 as shown on the right in the figure. The dominance of intentions directly determines embedding of segments.

One of the most important aspects of the work of G&S is the investigation of the relation between intentional and attentional structure in discourse. They model attentional state using a stack of focus spaces and a set of transition rules for pushing and popping spaces from the stack. Each focus space is associated with a discourse segment and contains the entities salient either because they have been mentioned explicitly in the segment or because they become salient in the process of comprehending the utterances in the segment. The DSP is also included in the focus space associated with a segment. G&S argue that attentional structure is parasitic on intentional structure; more precisely, the relationships among DSPs determine pushes and pops of focus spaces.

G&S provides a model of the use of referring expressions and aids in (a) determining the range of possible referents that an initial description makes available, and (b) choosing among the possibilities to determine the entity(ies) to which a pronoun or definite description refers. Determining how referring expressions interact with attentional state is crucial for solving these problems. A detailed account of the G&S approach to anaphora resolution is beyond the scope of this chapter. Instead we provide an example. Further detail can be found in Grosz and Sidner (1986) and Grosz et al. (1989).

Consider the example discourse in Fig. 12.8. G&S would break this discourse into two segments, DS_0 and DS_1, with the embedding shown in the figure. This structure accounts for why the utterances in 1–2 and 10–14 are more closely related to one another than to those in the intervening segment DS_1. In addition, G&S dictates that the focus space relating to the embedded segment DS_1 would be popped off the stack by the time the definite noun phrase *the tent* in 14 is interpreted, and thus explains how participant A can determine that B is referring back to the tent introduced in utterance 2 and not the tent introduced in utterance 7. Lexicogrammatical clues to this segmentation are given by *the last trip* and *this trip*.

As we have seen, DRT also deals with anaphora resolution, but DRT does not properly constrain the range of possible referents, and therefore both tents would be accessible. DRT overgenerates antecedents because it does not consider intention or the relation of intention to

DS0	1	A:	I'm going camping next weekend. Do you have a two person tent I could borrow?
	2	B:	Sure. I have a a two-person backpacking tent.
DS1	3	A:	The last trip I was on there was a huge storm.
	4		It poured for two hours.
	5		I had a tent, but I got soaked anyway.
	6	B:	What kind of a tent was it?
	7	A:	A tube tent.
	8	B:	Tube tents don't stand up well in a real storm.
	9	A:	True.
	10	B:	Where are you going on this trip?
	11	A:	Up in the Minarets.
	12	B:	Do you need any other equipment?
	13	A:	No.
	14	B:	Okay. I'll bring the tent in tomorrow.

FIG. 12.8. Discourse structure affects referent accessibility.

discourse segmentation and attentional structure, and therefore cannot rule out entities introduced in utterances 3–9. Also note that a full account of pronoun resolution requires a more fine-grained model of attentional state than that provided by attentional structure. An account of that phenomenon that is compatible with G&S is provided by centering theory, which is described in detail in Grosz et al. (1995). Finally, it has been recognized that there are significant difficulties in recognizing the speaker's discourse plan. These are thoroughly described in Grosz et al. (1989).

Rhetorical Structure Theory

As we have seen, in Grosz and Sidner's theory, speaker intentions and the relations of dominance and satisfaction precedence that may obtain among them are the primary determiners of discourse structure. Other researchers such as Hobbs (1983, 1985) downplay the importance of intention, arguing that the role of the speaker's intention is indirect and there are many discourse problems for which the speaker's intentions are uninformative.

The relation-based theories attempt to explain how coherent discourse conveys more than the sum of its parts. They aim to provide a more detailed account of the inferences that hearers can be expected to make when interpreting a series of utterances based on the assumption that they form a coherent discourse.

Hobbs (1979, 1983, 1985) characterized coherence in terms of a set of binary *coherence relations* between a current utterance and the pre-

ceding discourse. He identified four reasons why a speaker breaks a discourse into more than one clause and classified the relations accordingly. For example, if a speaker needs to connect new information with what is already known by the hearer, the speaker chooses one of the linkage relations, such as BACKGROUND or EXPLANATION. If a speaker wishes to move between specific and general statements, he or she must employ one of the expansion relations, such as ELABORATION or GENERALIZATION. According to Hobbs, how the speaker chooses to continue a discourse is equivalent to deciding which relation to employ. From the hearer's perspective, understanding why the speaker continued as he or she did is equivalent to determining what relation was used.

Hobbs (1979) originally proposed coherence relations as a way of solving some of the problems in interpreting discourse (e.g., anaphora resolution). He defined coherence relations in terms of inferences that can be drawn from the propositions asserted in the items being related. For example, Hobbs (1985) defined ELABORATION as follows:

> ELABORATION: S_1 is an ELABORATION of S_0 if the hearer can infer the same proposition P from the assertions of S_0 and S_1. (p. 25)

Here S_1 represents the current clause or larger segment of discourse and S_0 an immediately preceding segment. S_1 usually adds crucial information, but this is not part of the definition because Hobbs wishes to include pure repetitions under ELABORATION.

Hobbs' theory of coherence is attractive because it relates coherence relations to the functions that speakers wish to accomplish in a discourse. Thus, Hobbs' theory could potentially be used in text generation to indicate what coherence relation should be used to achieve a particular goal of the speaker. For example, Hobbs (1979) noted two functions of ELABORATION. One is to overcome misunderstanding or lack of understanding, and another is to "enrich the understanding of the listener by expressing the same thought from a different perspective." However, note that such specifications of the speaker's intentions are not an explicit part of the formal definition of the relation. For this reason, many researchers in text generation have turned to an alternative theory of text structure—Rhetorical Structure Theory (RST; Mann & Thompson, 1988).

As in Hobbs's account, RST characterizes coherence in terms of a set of relations between contiguous spans of text in a discourse. One of the main aims of RST was to account for conventional inferences that arise when interpreting discourse. For example, Mann and Thompson (1986) argued that in addition to the propositions represented explicitly by independent clauses in a text, there are many *implicit* propositions, called *relational propositions*, which arise out of the combination of clauses. They argued that the predicates of these propositions come from a small set of general, frequently occurring relational predicates (e.g., *cause*, *solutionhood*, *concession*). These propositions arise from the hearer's search for coherence among utterances that occur together—that is, from the hearer's assumption that the parts of a discourse form an intentionally constructed sequence of linguistic expressions. They emphasize that relational propositions are inferences that arise from the combination of the two parts of a text and cannot be derived from either of the parts independently.

Consider the following example from Mann and Thompson (1986):

(5) a. I'm hungry.
b. Let's go to the Fuji Gardens.

Mann and Thompson (1986) argued that the assumption that this is a coherent discourse gives rise to the relational proposition that (5b) provides a *solution* to the *problem* posed by (5a). The relational predicate associated with this proposition is called *solutionhood*. Note that although the solutionhood relation is not explicitly expressed anywhere in the text, it arises out of the juxtaposition of the two discourse elements. As we discuss later, it is crucial that systems generating such texts recognize these implicit relations that are conveyed because they may be the source of problems if the user does not understand or accept the system's utterance.

Mann and Thompson made an argument that relational propositions are more basic than other sorts of inferences that arise from texts and cited as evidence the fact that virtually every language has conjunction morphemes to signal them (e.g., in English, *because*, *therefore*, *so*, *however*). Rhetorical Structure Theory attempts to define a set of rhetorical relations that accounts for these relational propositions.

The definition of each *rhetorical relation* in RST indicates constraints on the two entities being related, constraints on their combination, as well as a specification of the effect that the speaker is attempting to achieve on the hearer's beliefs or inclinations. Thus, RST provides an explicit connection between the speaker's intention and the rhetorical means used to achieve it.

As an example, consider the RST definition of the MOTIVATION relation shown in Table 12.1. As shown, an RST relation has two parts: a *nucleus* (*N*) and a *satellite* (*S*). The MOTIVATION relation associates text expressing the speaker's desire that the hearer perform an action (the nucleus) with material intended to increase the hearer's desire to

TABLE 12.1
RST Relation—MOTIVATION

relation name:	MOTIVATION
constraints on N:	Presents an action (unrealized with respect to N) in which the Hearer is the actor.
constraints on S:	none
constraints on N + S combination:	Comprehending S increases the Hearer's desire to perform the action presented in N.
effect:	The Hearer's desire to perform the action presented in N is increased.

perform the action (the satellite). For example, in the following text, (6a) and (6b) are related by MOTIVATION:

(6) a. Come to the party for the new president.
b. There will be lots of good food.

The nucleus of the relation is that item in the pair that is most essential to the writer's purpose. In the prior example, assuming that the writer's intent is to make the hearer go to the party, clause (6a) is nuclear. In general, the nucleus could stand on its own, but the satellite would be considered a nonsequitur without its corresponding nucleus. In this example, without the recommendation to "come to the party," the satellite in (6b) is out of place. Moreover, RST states that the satellite portion of a text may be replaced without significantly altering the intended function of the text. The same is not true for the nucleus. For example, replacing (6b) with:

(7) *b* . All the important people will be there.

does not greatly change the function of the text as a whole. However, replacing the recommendation in the nucleus—for example,

(8) *a* . Don't go to the party.

significantly alters the purpose of the text.

Mann and Thompson also argued that if all the satellite units from a given text are deleted but all the nuclei remain, we should be left with a coherent text with a message resembling that of the original; it should be something like a synopsis or summary of the original text. As we see later, this observation has been useful in recent work on summarization.

In RST, schemas define the structural constituency arrangements of text. They are abstract patterns consisting of a small number of con-

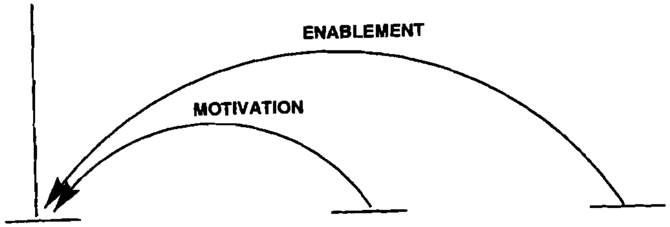


FIG. 12.9. Graphical representation of an RST schema.

stituent text spans, a specification of the relations between them, and a specification of how certain spans (nuclei) are related to the whole collection. Schemas are thus loosely analogous to grammar rules, and they constrain the possible RST structures. A graphical depiction of one schema defined by Mann and Thompson (1988) appears in Fig. 12.9. This schema consists of a nucleus and two satellites: one providing MOTIVATION for the material in the nucleus, and the other providing ENABLEMENT for the material in the nucleus.

RST schemas are recursive; text serving as a nucleus or satellite in one schema may itself be described by a schema that can be further decomposed into spans related in one of the ways dictated by the schemas. As defined by Mann and Thompson (1988), the schemas do not constrain the ordering of the nucleus and satellites, and each constituent relation may occur any number of times within a schema.

For example, the following text is described by the schema depicted in Fig. 12.9:

(9) a. Come to the party for the new president.
  b. There will be lots of good food.
  c. The Fluted Mushroom is doing the catering.
  d. The party is in the ballroom at eight o'clock on Friday.

In this example, (9a) is the nucleus of the entire text, and it presents an action that the speaker wishes the hearer to perform. (9b–c) presents information intended to increase the hearer's desire to perform the action, and is therefore a satellite related to (9a) by the MOTIVATION relation. (9b–c) is further decomposed into a nucleus, (9b), and a satellite, (9c), which in this case are related by EVIDENCE because (9c) is intended to increase the hearer's belief in (9b). In (9d), the speaker provides information intended to increase the hearer's ability to perform the action in the nucleus, and thus (9d) is a satellite span related to (9a) by the ENABLEMENT relation. The RST analysis of 9 is shown in Fig. 12.10.

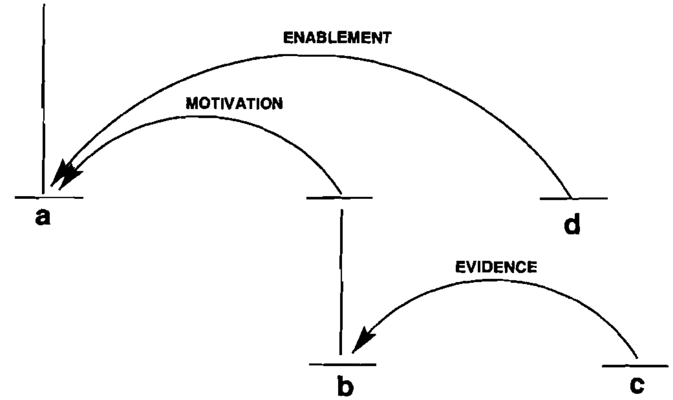


FIG. 12.10. Graphical representation of an RST analysis of (9).

To be acceptable, an RST analysis of a text must meet several additional criteria. It must be *complete* and *connected* (i.e., there must be one schema application under which the entire text is subsumed and all minimal units of the text must be accounted for in the analysis). In addition, each minimal unit can appear in exactly one schema application, and the spans constituting each schema application must be adjacent in the text. These constraints guarantee that a correct RST analysis forms a tree structure.

As we describe in more detail later, RST has been used extensively by researchers in text generation. More recently, RST has been used as the basis for *rhetorical parsing*, which has been applied to the problem of text summarization (Marcu, 1999). Although it did not correctly identify as many of the rhetorical relations as humans do (47% compared with 83%), the relations that it did identify were mostly correct (78% compared with 83% for humans; Marcu, 1999).

Segmented Discourse Representation Theory

The three theoretical approaches presented so far each focus on different aspects of what makes discourse coherent. Recently Asher and Lascarides developed a theory that combines the logic-based structures of DRT with the focus on rhetorical relations from RST to address a wide range of discourse phenomena. This theory, called Segmented Discourse Representation Theory (SDRT), started in Asher (1993) and has been further developed in Asher and Lascarides (1995) and Asher and Lascarides (1998, 2003).

In DRT, the discourse update procedure that joins the DRSs of two utterances together consists simply of appending the two structures.

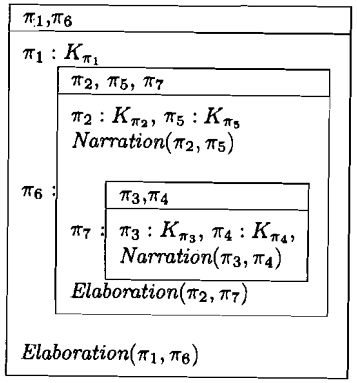


FIG. 12.11. An SDRS for discourse (10).

Any unresolved references to, for example, pronouns are resolved structurally (i.e., all available [not embedded] discourse referents are potential antecedents). Thus, DRT overgenerates antecedents; it allows coreferences that humans would never consider for a variety of discourse-related reasons. The connections are made on the basis of structure alone, not on the content.

SDRT greatly expands the power of the discourse update procedure by including rhetorical relations. Every time a DRS for a new utterance is added, some relation must be computed between it and one of the preceding utterances. The set of relations is open ended, but includes Narration, Elaboration, Continuation, Background, Explanation, Result, Evidence, Parallel, and Contrast. The relations are derived from theories in the field of pragmatics (e.g., Grice, 1957).

For each new utterance in a dialogue, a DRS is created in the same way as described previously. When it is added to the structure for the dialogue (the discourse update procedure), there must be some link established via a rhetorical relation with a preceding utterance. The inclusion of the relation constrains how the preceding discourse utterances can be accessed. Thus, the set of possible antecedents is not just based on structure, it is based on the pragmatically preferred reading.

For example, Fig. 12.11 shows an SDRS created from the discourse in (10), if K to K are DRSs that represent respectively the content of the utterances (10a) to (10e):

(10) a. Andrew's family had a great time at the beach last week. 1
b. They went snorkeling. 2
c. They saw a starfish. 3
d. They saw a grouper fish too. 4
e. Then they had dinner at a cafe on the beach. 5

The $_i$ symbols label subcomponents (representations of utterances) of the discourse. Relationships between them are given in the same way as normal DRT conditions—for example, *Narration*($_2$, $_5$) in this case. The key benefit of SDRT is that the specification of the relationships between the utterances constrains further processing—for example, anaphora resolution. If discourse (10) were continued with the sentence "It was delicious," the pronoun *it* could not refer to the grouper fish because its representation is embedded within a substructure.

We know of no current implementation of SDRT. It was previously used in a project for analyzing air traffic control conversations in French (Asher, Aurnague, Bras, & Vieu, 1995). It is the center of some recent research in natural language generation (Danlos, Gaiffe, & Roussarie, 2001) and in dealing with sentence fragments in discourse (Schlangen, Lascarides, & Copestake, 2001).

GENERATING COHERENT DISCOURSE

As noted earlier, a comprehensive survey of discourse structure for natural language understanding appears in Grosz et al. (1989). Thus, here we focus on the role of discourse in natural language generation.

Content Selection and Organization

Early work in natural language generation (Appelt, 1985; Cohen, 1978; Cohen & Perrault, 1979) focused on generating utterances that would allow a hearer to recognize the speaker's intention to perform a particular speech act. These systems formalize the preconditions and effects of illocutionary acts, and reason about the beliefs of the hearer and speaker and the effects of speech acts on these beliefs. This approach explicitly represents the relation between the speaker's intentions and the speech acts that achieve them. This is a necessary component of any system that must participate in a dialogue with its users. However, these systems could generate only short (one- or two-clause) texts; they do not represent or use knowledge about how speech acts may be combined into larger bodies of coherent text to achieve a speaker's goals.

To build systems capable of producing larger bodies of text, researchers sought approaches that would enable systems to produce texts that adhere to standard patterns of discourse and flow smoothly from topic to topic. Until recently, two main approaches to the generation of connected discourse were prevalent: graph traversal and schema based. Graph traversal produces texts whose structure mirrors the structure of the domain knowledge representations being traversed, whereas schema-based approaches use domain-independent rhetorical strategies to select information and impose a structure on the text. Both approaches incorporate focus considerations to determine what to say next when several options are available.

Graph Traversal: Paraphrasing the Knowledge Base

By studying a range of naturally occurring texts, computational linguists identified types of text for which discourse structure closely follows the conceptual relations (e.g., causal, temporal, spatial) connecting the domain entities being described. For example, Paris (1988) observed that one strategy for describing a complex physical object is to trace through the process that allows the object to perform its function. Sibun (1992) found that texts describing the layout of houses or the relations among members of a family also follow domain relations. When text structure follows domain structure, the most appropriate generation mechanism selectively traverses existing links in a knowledge base (Suthers, 1991). To generate text, graph traversal is typically combined with a focusing mechanism, which appeals to some model of attentional structure to select the most appropriate thing to say next when multiple domain links are available.

Uses of this technique are limited to cases where the system's representation of the domain mirrors or can be engineered to mirror the structure of natural utterances in the domain. This requirement may place an undue burden on knowledge engineers, who are already trying to mediate between the sometimes conflicting demands of writing programs that are correct, efficient, and maintainable (Swartout, 1983). Moreover, the approach does not admit communicative strategies that depart from the organization of the knowledge base, thereby restricting the types of texts that a system can produce. Finally, because the approach does not model the effects of its utterances, it does not provide a representation of the text from which a system can determine how to interpret and respond to feedback from the user.

Schemas: Representing Patterns of Rhetorical Structure

Some of the limitations of traversal techniques can be overcome by using domain-independent rhetorical strategies that control both what is said and how it is structured. Many researchers have attempted to understand the nature of the relationships that hold between the utterances of a coherent discourse (Lehnert, 1981; Polanyi, 1988; Reichman-Adar, 1984; Rumelhart, 1975). At the heart of many of these efforts is a set of rhetorical predicates or relations that characterize the organization of coherent texts of the type studied. We have seen that many linguists and computational linguists have proposed classifications of these relations and attempted to identify their functions (Hobbs, 1983; Grimes, 1975; Lascarides & Asher, 1993; Longacre, 1983; Mann & Thompson, 1988).

McKeown (1985) found a correlation between the discourse purpose of a text and patterns of rhetorical predicates in the text, with a family of similar patterns representing a *strategy* for achieving a given discourse purpose. McKeown encoded these patterns of rhetorical predicates into script-like (Schank & Abelson, 1977) structures called *schemas*. By associating each rhetorical predicate with an access function for an underlying knowledge base, these schemas can be used to guide both the selection of content and its organization into a coherent text to achieve a given communicative goal. The schema-based and other approaches based on rhetorical structuring (Hovy, 1988b) have proved successful for many text generation applications.

The graph traversal and rhetorical structuring approaches to content determination are not mutually exclusive. For example, the Intelligent Labelling EXplorer (ILEX; Mellish et al., 1998) uses a combination of these techniques. In ILEX, facts about the domain are organized into a *text potential*—a graph of facts interconnected in terms of thematic and rhetorical relations representing the information ILEX can express and the ways it can be related. Given an entity to describe, ILEX finds all of the facts associated with that entity and selects among these opportunities for what to say next using a set of heuristics. Once a fact is chosen, all of its connections are examined, the next fact is chosen, and so on.

The schema-based approach has several advantages. First, because it decouples discourse strategies from low-level details of knowledge representation, knowledge engineers have more flexibility to design knowledge bases to satisfy other desiderata, such as maintainability and run-time efficiency. Second, discourse strategies based on rhetorical knowledge enable systems to generate a range of different texts from the same knowledge representation.

However, Elhadad (1996, 1997) has argued the need to go beyond schemas to produce effective argumentation. Moore (1995) has shown that schemas and rhetorical structure trees are insufficient as a discourse model for dialogue systems because they do not include a representation of the intended effects of the components of the text produced, nor how these intentions are related to one another or to the informational structure of the utterances produced. A schema can be viewed as the result of a compilation process that dispenses with the rationale for all of its component steps. What remains is the top-level communicative goal that invoked the schema and the sequence of actions (i.e., instantiated rhetorical predicates that cause sentences to be generated) that are used to achieve that goal. Because of this compilation, schemata provide a computationally efficient way to produce multisentential texts for achieving discourse purposes. They are rhetorical recipes that encode frequently occurring patterns of discourse structure. Using schemata, the system need not reason directly about how speech acts affect the beliefs of the hearer and speaker, nor about the effects of juxtaposing speech acts. The system is guaranteed that each schema will lead to a coherent text that achieves the specified discourse purpose.

However, this compilation renders the system incapable of responding appropriately if the hearer does not understand or accept the utterance. Because intentional structure has been compiled out of the schema representation, the system cannot determine whether any of the discourse actions in the recipe have failed to achieve their intended effects or what other strategies could be used to achieve those effects. Intentional structure is crucial for interpreting and responding to questions or challenges that address a previous utterance: Without a record of what an utterance was intended to achieve, it is impossible to elaborate, clarify, or defend that utterance. This is because there is not a one-to-one mapping between intentional and informational structure. That is, there is not a one-to-one mapping between the ways in which content can be related in a discourse and the ways in which intentions combine into a coherent discourse plan to affect a hearer's mental state (Moore & Paris, 1993; Moore & Pollack, 1992). Thus, it is impossible to recover intentional structure from informational structure or vice versa. Therefore, it is not possible to reconstruct the intended effects of individual actions in an instantiated schema, which contains only propositions and rhetorical relations between those propositions.

Plan-Based Approaches

To overcome the limitations inherent in schema-based approaches, researchers have applied techniques from AI planning research to the problem of constructing discourse plans that explicitly link communicative intentions with communicative actions and the information that can be used in their achievement (Moore, 1995; Young et al., 1994a). Text planning generally makes use of *plan operators*—discourse ac-

tion descriptions that encode knowledge about the ways in which information can be combined to achieve communicative intentions:

- effect(s): communicative goal(s) the operator is intended to achieve.
- preconditions: conditions that must hold for an act to successfully execute. For example, it may be the case that the hearer must hold certain beliefs or have certain goals for a particular discourse strategy to be effective.
- constraints: specifications of the knowledge resources needed by the discourse strategy.
- subplan: optionally, a sequence of steps that implement the discourse strategy.

Simplified examples of typical discourse planning operators, taken from Young and Moore (1994), are shown in Table 12.2. In this framework, the representation of communicative action is separated into two types of operators: action and decomposition. Action operators capture the conditions (preconditions and constraints) under which an action can be executed and the effects the action achieves if executed under the appropriate conditions. Preconditions specify conditions that the agent should plan to achieve (e.g., the hearer knows a certain term), whereas constraints specify conditions that the agent should not attempt to plan to change (e.g., facts and rules about the domain). Effects describe the changes that a discourse action is intended to have on the hearer's mental state. If an action is composite, there must be at least one decomposition operator indicating how to break the action down into more primitive steps. Each decomposition operator pro-

TABLE 12.2
Discourse Plan Operators

Operator 1: Action operator for Cause-to-Bel
HEADER: Cause-to-Bel(?p)
CONSTRAINTS: not(Bel(?p))
PRECONDITIONS: nil
EFFECTS: Bel(?p)
Operator 2: Decomposition operator for Cause-to-Bel
HEADER: Cause-to-Bel(?p)
CONSTRAINTS: nil
STEPS: Begin, Inform(?p), Support(?p), End
Operator 3: Decomposition operator for Support
HEADER: Support(?p)
CONSTRAINTS: causes(?q, ?p)
STEPS: Begin, Cause-to-Bel(?q), Cause-to-Bel(causes(?q, ?p)), End

vides a partial specification for a subplan that can achieve the action's effects, provided the preconditions are true at the time the steps in the decomposition are executed.

As an example of how action and decomposition operators are used to encode discourse actions, consider operators 1 and 2 in Table 12.2. These two operators describe the discourse action Cause-to-Bel, which is capable of making the hearer believe a proposition. Operator 1 is an action operator; it indicates that Cause-to-Bel can be used to achieve the state where the hearer believes a proposition ?p, if the hearer does not already believe ?p.[3] Operator 2 in Table 12.2 is one decomposition operator for the Cause-to-Bel action. It says that one plan for making a hearer believe a proposition is to inform the hearer of that proposition and provide support for that proposition.

As illustrated in operator 3 of Table 12.2, decomposition operators may also have constraints, which indicate the conditions under which the decomposition may be applied. Such constraints often specify the type of information needed for particular communicative strategies, and satisfying them causes the planner to find content to be included in its utterances. For example, operator 3 encodes the discourse strategy that one way for a speaker to support a proposition is to describe a plausible cause of that proposition. More specifically, the constraint on operator 3 says that to support proposition ?p, there must be another proposition, ?q, such that causes(?q, ?p) is true in the domain. When the planner attempts to use a decomposition operator to support a proposition, it must try to satisfy all of its constraints. If a constraint contains no unbound variables, it is simply checked against the knowledge source to which it refers. However, if the constraint contains free variables (e.g., ?q in operator 3), the system must search its knowledge bases for acceptable bindings for these variables. In this way, satisfying constraints directs the planner to select appropriate content to include in explanations. In operator 3, if an appropriate ?q can be found, then the speaker can support ?p by making the hearer believe ?q and making the hearer believe that causes(?q, ?p). Thus, we see that action and decomposition operators specify how information can be combined in a discourse to achieve effects on the hearer's mental state. That is, action operators and their decompositions encode the link between intentional and informational structure.

A detailed description of the algorithm for synthesizing plans from such operators is beyond the scope of this chapter and may be found in (Young & Moore, 1994; Young, Pollack, & Moore, 1994b; Moore, 1995). However, it is important to note that the plans produced from such operators can serve as the speaker's model of the effects that individual parts of the text were intended to have on the hearer and how they fit together to achieve her top-level intention. From a text plan, we can determine which intentions dominate other intentions (i.e., we can determine when an intended action is in the plan to serve a higher intention). This allows the speaker to interpret and respond to feedback indicating that the hearer does not fully understand or accept what the speaker says and localize the failure in some portion of the utterance that failed to achieve its intended purpose.

[3]Note that not(Bel(?p)) is a constraint because we do not wish our discourse planner to attempt to plan to make the hearer not believe ?p to use this operator.

TABLE 12.3
Fragment of a Naturally Occurring Tutoring Dialogue

TUTOR	< — > Next, you replaced the A1A3A13.	[1]
STUDENT	This is the first circuit card assembly in the drawer that the signal goes through. Why would I not start at the entrance to the station and follow the path to the measurement device?	[2]
TUTOR	It would have been better to troubleshoot this card by taking measurements instead of swapping. You can't count on having spare cards available for swapping.	[3]

To illustrate this, we consider an utterance from a naturally occurring tutorial dialogue and see how reasoning about the text plan that produced the utterance could enable a system to respond appropriately to a range of student responses. Consider the dialogue fragment shown in Table 12.3, which was taken from a corpus of student–tutor interactions in which students are using a computer simulation environment that trains them to troubleshoot the complex electronic circuitry found in aircraft. The tutoring system presents the student with a troubleshooting problem to solve, allows the student to solve the problem with minimal tutor interaction, and then provides a critique of the student's solution in a postproblem review session. During the review session, the system replays the student's solution step by step, pointing out good aspects of the student's solution as well as ways in which the student's solution could be improved.

In turn 1 of the dialogue in Table 12.3, the tutor indicates that the problem-solving step of replacing a particular component was suboptimal (as indicated by < — > preceding the step description). The student does not immediately accept this assessment and probes the tutor to find out why this action was assessed negatively, turn 2.

In turn 3, the tutor explains why the student's action was judged suboptimal, with the following utterance, repeated here for convenience:

(11) P1. It would have been better to troubleshoot this card by taking measurements instead of swapping.
P2. You can't count on having spare cards available for swapping.

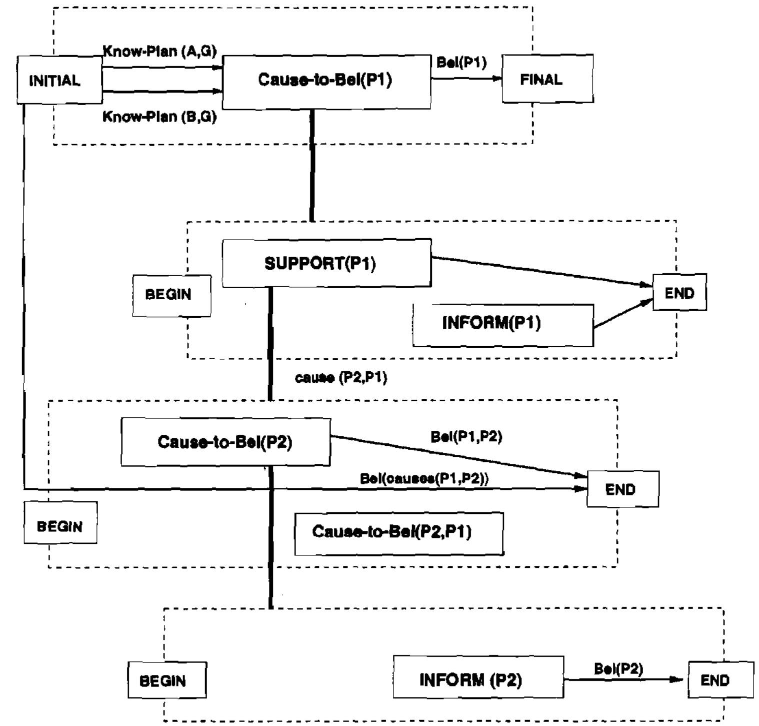
Our analysis of this example is that the tutor's primary intention is to convince the student of P1—that it is better to troubleshoot a component by taking measurements than to swap the component. To achieve this goal, the tutor asserts this proposition and then supports it by claiming P2—that the student cannot always count on having spare parts available for swapping. The plan for generating this text is shown in Fig. 12.12.

To handle feedback indicating that the hearer does not fully understand or agree with what the speaker says, the speaker must be able to determine what portion of the utterance failed to achieve its intended purpose. Therefore, the speaker must have a model of the effects that individual parts of the text were intended to have on the hearer and how they fit together to achieve the speaker's top-level intention. From the plan in Fig. 12.12, we can determine that the speaker's intention to make the hearer believe that one cannot count on having spare cards available for swapping serves the higher level intention of making the hearer believe that it would have been preferable to troubleshoot the card by taking measurements.

Now let us consider several possible student responses to the tutor's utterance in turn 3 of the sample dialogue, given in examples (12)–(14), and see how a computer system, acting as tutor, could use this discourse plan to help determine appropriate responses in each case.

(12) We always have spare cards in our shop.
(13) Yeah, it would have been better to troubleshoot the card, but we always have spare cards in our shop.
(14) Yeah, you can't count on having spares, but it's still better to swap.

In example (12), the student rejects P2—the proposition that spare cards may be unavailable. This blocks the support that P2 would have provided to convince the hearer that troubleshooting by taking measurements is a better strategy than swapping. To see how the system can determine this, consider again the discourse plan in Fig. 12.12. Response (12) indicates that the effect Bel(P2) asserted by the Inform(P2) was not achieved. From the plan representation, it is possible to trace a path of failed effects from Bel(P2) across causal links and up decompositional links through the actions Inform(P2), Cause-to-



P1 = It would have been better to troubleshoot by taking measurements instead of swapping.

P2 = You can't count on having spare cards available for swapping.

G = to troubleshoot the test station

A = taking measurements

B = swapping components

FIG. 12.12. A sample discourse plan.

Bel(P2), Support(P2), Cause-to-Bel(P1), and eventually to the top-level intended effect Bel(P1). Using this information, the system can determine that appropriate responses to (12) can be generated by trying to convince the student that spare cards are not, in fact, always available (i.e., replanning the subtree rooted at the node Cause-to-Bel(P2), most likely by providing support for P2) or by finding some other support for the claim that troubleshooting by taking measurements is a better strategy than swapping (i.e., replanning the subtree rooted at the node Support(P1)).

An appropriate response to (13) would be different. In (13), the hearer again expresses disbelief in the supporting proposition P2 (i.e., the effect Bel(P2) asserted by the Inform(P2) was not achieved). However, here the student gives more information about the success of the speaker's original plan by indicating that he believes it would have been better to take measurements than to swap—that is, here the effect Bel(P1) has been achieved. As in (12), the speaker's intention to get the hearer to believe P2 has failed, and thus the support that P2 would have provided for P1 is again blocked. However, in (13), the tutor need not do anything about this failure. This is because the intention to get the hearer to believe P2 was held in service of the intention to get the hearer to believe P1. Since (13) explicitly indicates that Cause-to-Bel(P1) has achieved its intended effect—namely, Bel(P1)—the outcome of any intended act that served as a precondition to this intention or as a step in a subplan for this intention can be ignored—unless of course the tutor has some other reason for wanting the student to believe P2.

Now consider how to handle (14), where the hearer agrees with P2, but is not convinced of P1. Implicit in the speaker's original argument was his belief that, as a rule, not being able to count on having spare cards makes troubleshooting by taking measurements a preferable strategy to swapping. That is, the discourse plan in Fig. 12.12 is predicated on the truth of Bel(causes(P2,P1)). Note that the node Cause-to-Bel(causes(P2,P1)) was not expanded in this discourse plan because this proposition was an effect of the initial step. Together with the statement in P2, causes(P2,P1) would have provided support to convince the hearer that it is better to troubleshoot before swapping. (14) indicates that the support for P1 has failed. At this point, the tutor must either convince the student that the causal relationship between P2 and P1 does indeed hold or must find another way to support P1. As in the previous case, an appropriate response results from the replanning of subtrees whose execution is affected by this failure. Specifically, those subtrees rooted at Cause-to-Bel(causes(P2,P1)) across causal links and up decompositional links to Support(P1) and eventually to Cause-to-Bel(P1). Note that this does not include the subtree rooted at Cause-to-Bel(P2) and thus, unlike in (12), the system will behave correctly and will not attempt to reestablish P2—the proposition that spare cards may not always be available.

In these examples, each of the hearer's replies provides feedback about a subset of the speaker's intentions. To respond appropriately, the speaker must reason about the relationships between the intentions in his communicative plan to determine what implication the hearer's feedback has on the ultimate success of his other intentions. These examples show that the information in discourse plans provides guidance for the planning of subsequent responses.

Note that discourse plans in this framework also include a representation of informational structure. When a discourse strategy requires that a particular informational relation be recognized by the hearer for a discourse to have its intended purpose, a condition expressing this requirement appears in the (sub)plan that requires this. For example, in Fig. 12.12, the strategy for supporting P1 is successful only if the two effects Bel(P2) and Bel(causes(P2,P1)) are achieved. Conditions such as this one allow the planner to recognize how the content expressed in portions of the text plan is related. Among other tasks, the informational structure is used by the realization component when transforming the discourse plan into a series of natural language utterances. The representation of informational structure, together with information about intentional structure, allows the text generator to order clauses and select appropriate content-bearing discourse cues (e.g., *because*, *in addition*).

Discourse plans such as these, which have been used in many systems, are hierarchical structures whose leaves specify a partially ordered sequence of speech acts to be performed. The internal structure of the plan represents dominance and satisfaction precedence relations between discourse intentions, and steps in subplans post goals to make the hearer recognize informational relations between plan components. Although these plan structures contain much information that is crucial for generating coherent multisential natural language texts, they cannot be transformed directly into natural language. They do not include all of the information required by existing syntactic realizers, which transform abstract syntactic specifications of natural language sentences (or phrases) into their corresponding surface forms. Examples of such systems are FUF (Elhadad, 1992) and REALPRO (Lavoie & Rambow, 1997).

To construct specifications from which English sentences can be generated, many decisions about organization and lexicalization remain. A system must choose a total ordering for the steps in the discourse plan and decide how to apportion propositional content to sentences and sentence constituents (Hovy, 1988a; McDonald, 1983; Meteer, 1992). It also must choose referring expressions (Appelt, 1985; Dale, 1992; Reiter, 1990) and lexical items to express the content in the speech acts (Bateman & Paris, 1989; Danlos, 1984; Granville, 1984; Matthiessen, 1991; McDonald, 1991; Pustejovsky & Nirenburg, 1987; Reiter, 1991). As the references indicate, there has been considerable research on many of these issues.

A remaining issue, and one that has received relatively little attention in the computational generation literature, concerns the use of discourse cues. *Discourse cues* are words or phrases, such as *be-*

cause, *first*, *although*, and *also*, that mark structural and semantic relationships between discourse entities. They play a crucial role in many discourse processing tasks, including plan recognition (Litman & Allen, 1987), text comprehension (Cohen, 1984; Hobbs, 1985; Mann & Thompson, 1986; Reichman-Adar, 1984), and anaphora resolution (Grosz & Sidner, 1986). Moreover, research in reading comprehension indicates that felicitous use of cues improves comprehension and recall (Goldman, 1988), but that indiscriminate use of semantic cues may have detrimental effects on recall (Millis, Graesser, & Haberlandt, 1993). In addition, there is evidence that the benefit of discourse cues may depend on the subjects' reading skill or level of prior knowledge of the domain (Goldman & Murray, 1992; Meyer, Brandt, & Bluth, 1980; McNamara, Kintsch, Songer, & Kintsch, 1996).

The problems of determining when discourse cues should be used in the final text, where the cues should be placed, and which cues would be most effective in increasing the hearer's comprehension of a text are a current area of research. McKeown and Elhadad studied several connectives (e.g., *but*, *since*, *because*) with the aim of identifying features of the propositions connected by the cues that can be used to select appropriate cues during text generation (Elhadad & McKeown, 1990; McKeown & Elhadad, 1991). Researchers concerned with generating text from RST trees (where leaf nodes contain content and internal nodes indicate the RST relation that holds between subtrees) have proposed algorithms for determining sentence boundaries and choosing cues based on the rhetorical relation between spans of text, the order of the relata, and the complexity of the related text spans (Rösner & Stede, 1992; Scott & de Souza, 1990).

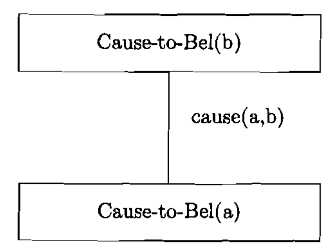
As noted earlier, RST analyses presume that there is a primary rhetorical relation between any two consecutive elements of a coherent discourse, and RST analyses do not permit multiple relations between text spans. This means that consecutive elements in RST trees are related *either* by an informational *or* an intentional relation. However, Moore and Pollack (1992) showed that discourse interpretation and generation cannot depend on analyses in which informational and intentional structure are in competition; intentional and informational analyses must coexist. Therefore, we must devise algorithms for generating appropriate texts from a discourse model that represents these two types of structure, such as the discourse plans described earlier.

Intentional and Informational Structure and Cue Choice

It should be clear that the informational (or semantic) relation between discourse items affects the choice of cue. For example, to mark an exemplification relation, a speaker can felicitously use a cue such as *for example* or *for instance*, but could not use a cue such as *thus* or *first*. It is less clear to many how intentional structure affects cue usage. Note here that we are concerned with those cues that convey semantic information between discourse elements, such as *because*, *thus*, or *for example*. It *is* clear that intentional structure affects cues that indicate purely structural aspects of discourse, such as topic shifts (e.g., *now*, *anyway*) and digressions (*by the way*; Grosz & Sidner, 1986; Hirschberg & Litman, 1993).

To illustrate the effect of intentional structure on cue choice, let us consider the two example discourses in (15) and (16) in which the informational relation between discourse entities and the placement of the cue is held constant, but the intentional structure varies. Figure 12.13 shows the intentional and informational relations between two discourse actions and the text that these actions produce. In this example, the tutor is trying to convince the student of (15b)—that there is a break in a certain signal path. To achieve this goal, the tutor informs the student of (15b) and supports it with (15a). In the domain, there is a causal connection between (15a) and (15b), the bad signal at pin 26 causes there to be a break in the particular path discussed. Thus, the tutor can use the discourse strategy (encoded in operator 3) of supporting a proposition by describing a plausible cause of that proposition. Figure 12.13 represents a typical deductive argument; to convince the hearer of an effect (15b), cite its cause (15a) as support.

In example 15, the intention to make the student believe (15b) dominates the intention to make the hearer believe (15a). At the informational level, there is a causal relation between (15a) and (15b). In the text, the proposition expressing the cause (and the dominated discourse purpose) precedes the one expressing the effect (and the dominating discourse purpose). The discourse cue, *thus*, is placed with the



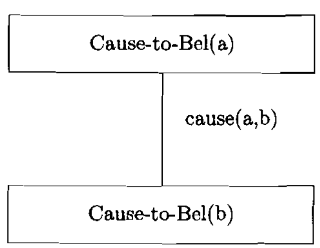
(15) a. You know that the signal on pin 26 is bad.
b. **Thus,** there's a break in the path created by TPA63.

FIG. 12.13. Arguing from cause to effect.

proposition expressing the effect. This cue indicates both the causal relation at the informational level as well as the dominance relation between the speaker's intentions.

In contrast, consider the relation between the discourse actions and the accompanying text in Fig. 12.14. In this example, the tutor is trying to convince the student of (16a)—that the signal on pin 26 is bad and is using (16b) as support. That is, the speaker is trying to convince the hearer that a state exists by citing an effect of that state. This is a typical abductive argument. In this example, the informational relation between (16a) and (16b) is the same as in example (15)—that is, (16a) causes (16b). However, the two texts differ at the intentional level. In (16), the intention to make the hearer believe (16a) dominates the intention to make the hearer believe (16b). This difference in intentional structure is reflected in the discourse cue chosen. As in example (15), the cause precedes the effect, and the discourse cue is placed with the text expressing the effect. However, a different discourse cue (*because*) must be used to indicate the difference in intentional structure. In (16), the intentional roles of cause and effect are reversed. The proposition expressing the cause is now expressing the dominating discourse purpose, and the one expressing the effect is now expressing the dominated discourse purpose. The cue is now placed with the proposition expressing the dominated discourse purpose. Since the causal relation at the informational level has remained unchanged from example (15), the difference in cue must be due to the difference in intentional structure.

These examples show that algorithms based on a discourse model that forces a choice between intentional and informational structure, such as RST, cannot be complete. Algorithms for cue usage must take



(16) a. You know that the signal on pin 26 is bad
b. **because** there's a break in the path created by TPA63.

FIG. 12.14. Arguing from effect to cause.

both informational and intentional structure into account. Moser and Moore (1995, forthcoming) have done an extensive empirical study to identify the factors that predict appropriate use of cues. Di Eugenio, Moore, and Paolucci (1997) are using machine learning to induce decision trees that can be used to determine cue choice in automatic text generation.

CURRENT DISCOURSE APPLICATIONS

In this section, we describe some of the new directions in computational linguistics and AI in processing discourse. Many of the current techniques break with the theoretical traditions described in previous sections. Instead, they rely on shallow text processing techniques and statistical methods that support the inference of discourse information in a task-specific or domain-specific way.

Summarization

As its name suggests, the goal of a text summarization system is to produce a summary of a text that can be quickly read by a user. Because of the huge amounts of textual data available on the Web and elsewhere, text summarization can provide a great benefit to those who need to scan or stay current in a topic, but care only about the main points and not the details.

Most current systems perform summarization in three steps:

1. Identify the important text units of a documents (or set of related documents).
2. Extract from each unit the most central sentence or sentences.
3. Join them to form the summary.

This section describes how discourse processing techniques are used to perform these steps in a variety of summarization systems.

A critical first step in determining which parts of a document are most important is locating the boundaries between topic segments. This allows a summarization system to know which topics should be represented in the summary, and prevents the system from misreading important text on a new topic as less important text continuing the current topic.

The standard technique for performing automatic text segmentation is to use some measure of sentence similarity to find consecutive clusters of sentence that have something in common. The usual similarity metrics are based on word overlap, typically by calculating word stems, and then comparing sentences with either a keyword overlap mechanism or vector-based formula. Then some type of clustering algorithm is used to detect boundaries between groups of consecutive sentences that are highly similar (Choi et al., 2001; Hearst, 1997).

Another approach is to use the pattern of word occurrences across the text as an indication of topic segments. Lexical chaining was introduced by Morris and Hirst (1991). For each reference to an entity in the text, a graph is created that follows references to the same or closely related terms. The related terms are inferred from synonym sets provided by a lexical database like WordNet (Fellbaum, 1998). The chains show which terms are essential to the text as a whole (longer chains represent more significant terms) and locate where those terms are mentioned (Barzilay & Elhadad, 1997; Hirst & St-Onge, 1998; Stairmand, 1996).

A genre-specific segmentation method was developed by Teufel (1999). For scientific papers, she used a machine learning technique to associate a variety of discourse cues with argumentative zones—that is, sections of the papers with different functional purposes. Two examples of argumentative zones are general scientific background and descriptions of other people's work. These zones provided the basis for performing topic segmentation of the texts.

Marcu (2000) used a similar technique that works on unrestricted texts. His approach, rhetorical parsing, also used machine learning to determine the rhetorical relations based on a variety of shallow cues such as discourse cues and punctuation. For example, the use of the cue term *although* indicates either a Concession or an Elaboration relation with a neighboring textual unit. Marcu demonstrated that the hierarchical rhetorical trees that this technique produces are useful for text summarization because they highlight topic switches in the text. The rhetorical relations also indicate the central text units of each section. Systems that do not perform rhetorical parsing tend to rely on word overlap measures to determine which text units are most central within a section.

The key sentences of a segment cannot normally be used directly in a summary because of anaphoric references. Coreference resolution must be performed to replace abbreviated references with their fully specified form. For example, a document might refer to the same company as "Apple Computer Inc," "Apple," "the computer company," "the company," and "it." Because the first reference may not be in the sentence that contains the most central information, coference resolution is essential. As mentioned earlier, the various computational theories of discourse structure each have something to say about the constraints on coreference. Because most summarization systems do not perform in-depth processing of the texts, they tend to use domain-specific shallow methods to track coreference.

Coreference resolution has been the focus of many current corpus-based applications. Lappin and Leass (1994) described a model for resolving pronominal anaphora using a surface parse of the sentences and a heuristic measure of salience of potential referents based on features such as their proximity and syntactic position. This approach was extended by Kennedy and Boguraev (1996) to use an even shallower syntactic representation in which words were marked with their syntactic function, but no parse tree was required. Other researchers have used machine learning approaches like Bayesian induction and decision trees to learn methods for coreference resolution from annotated corpora (Aone & Bennett, 1995; Connelly, Burger, & Day, 1994; Kehler, 2000; McCarthy & Lehnert, 1995). A recent DARPA-sponsored information extraction initiative (Sundheim, 1995) had a subtask that required participants to resolve coreference among proper names, aliases, definite noun phrases, and more.

As in other natural language processing tasks, performance is normally calculated by comparing with human judgments and reported in terms of *recall*, *precision*, and *Fscore*. Recall is the number of true positives divided by the sum of the true positives and the false negatives. Precision is the number of true positives divided by the sum of the true positives and the false positives. Because there is normally a trade-off between precision and recall, the Fscore combines them and is defined as: 2 * *Precision* * *Recall*/(*Precision* + *Recall*). On the coreference task, participants have achieved success rates of over 60% recall, 70% precision, and 65% Fscore (Sundheim, 1995).

Once coreference resolution is performed by a summarization system, the fully specified sentences can then be processed into a coherent and readable summary. Using techniques from natural language generation, references to the same entities can be replaced with anaphoric expressions to increase the cohesiveness of the text. Sentences may also need to be restructured to account for discourse focus (Kan & McKeown, 1999).

Another DARPA initiative has recently compared the performance of several text summarization systems (Mani et al., 1998). Because it is so difficult to determine what an ideal summary would be, the evaluation focused on extrinsic metrics—that is, ones that measure how well the end product of the system enables a human to perform some task. In this case, one task (the ad hoc task) was to determine whether a document was relevant to a particular topic. A good summary allows the human analyst to quickly determine whether the source document is relevant. In this evaluation, the best systems achieved performance of over 60% recall, 80% precision, and Fscores of around 70%. There was also a categorization task, in which the summaries were used by a human analyst to classify the source document as either fitting into one of five topics or *none of the above*. Here the scores were lower, with top precision scores of around 70%, recall scores around 45%, and Fscores around 50%.

Question Answering

Question answering (QA) is an offshoot of the information retrieval task. In information retrieval, the task is to select from a large database of texts a small number that matches some query, similar to exploring the Internet with a search engine. In QA, the task is more specific: Instead of returning an entire document, the system should return just the specific sentence that answers the question.

The standard approach to this task involves first performing information retrieval to find relevant documents. Then each document is searched for the sentence(s) that are most relevant to the question. The sentences are ranked for relevance and informativeness, and the highest ranking sentence is returned as the answer.

Here also it is clear that discourse information plays a crucial role. For example, in searching for the answer to the question, "Who wrote "To Kill a Mockingbird"?, a system might find the text:

> Now Harper Lee is 70, a white-haired woman who protects her privacy as insistently as J.D. Salinger or Thomas Pynchon. She is the author of a single book, "To Kill a Mockingbird," but that book is one of the most widely read American novels of the century.

A QA system must resolve the coreference to succinctly provide the correct answer. Coreference resolution has been used to increase the performance of a number of recent QA systems (Breck, Burger, Ferro, House, Light, & Mani, 1999; Humphreys, Gaizauskas, Hepple, & Sanderson, 1999; Litkowski, 1999). The systems used a variety of more local techniques, for example, producing variants of the questions. Because these are not discourse related, they are not described here.

In the most recent QA section of the DARPA-sponsored TREC (Text REtrieval Conference) program, the task was to answer a set of approximately 700 fact-based short answer questions by extracting for each a small amount of text (250 bytes) from a 3-gigabyte collection of newswire text. An example question is, "How much folic acid should an expectant mother get daily?" The systems were allowed to provide a ranked set of answers to each question. The scores were based on how far down the stack of answers the correct answer was or 0 for no correct answer. The best system from among the 28 participants achieved a score of 75% (the correct answer was on average one quarter down the ranked list of answers) and did not answer 15% of the questions.

SUMMARY

In this chapter, we have discussed the types of discourse structure that researchers in computational linguistics and AI have developed to address a range of problems in discourse interpretation and generation. In conclusion, we would like to point to several fruitful areas for future research.

First, discourse actions like all other actions have context-dependent effects. Indeed, one of the main aims of discourse research is to understand how the context in which an utterance occurs affects the meaning of that utterance. If we are taking a plan-based view of discourse interpretation and generation, many inferences beyond what are listed as the direct effects of discourse operators may be licensed. In AI, this is the well-known *ramification problem*. In discourse, some of these inferences, the Gricean conversational implicatures, are an important part of normal cooperative conversation. As we have argued, discourse interpreters must make these inferences if they are to properly understand the discourse, and discourse generators must be aware of the implicatures licensed by their utterances to produce natural sounding utterances and avoid leading the hearer to make false implicatures. Although early research (e.g., Hirschberg, 1985; Joshi, Webber, & Weischedel, 1984) identified the problem and attempted to specify the conversational implicatures licensed by certain types of utterances, there has been no general solution. Recently, there has been renewed interest in this problem, and considerable progress has been made (see Green & Carberry, 1999; Stone, 2001). However, much work remains to be done to provide an adequate model of discourse actions and their effects, which can be used in realistic computational systems.

Second, computational accounts of discourse have largely ignored the impact that human processing limitations must have on language. However, some research has shown that taking resource limitations seriously can provide an explanation of phenomena such as how much risk speakers take of being misunderstood in a given conversational setting (Carletta, 1992) and why speakers sometimes produce redundant utterances (Walker, 1996). Moreover, McKeown's (1985) schema-based generation system showed how constraints on focus of attention could affect the content and organization of a discourse. To adequately model human discourse interpretation and generation in computational systems, we must further investigate the interpretation and generation problems as resource-bounded tasks.

Third, we must take seriously a model of belief and context update. Without it, our theories cannot adequately account for why speakers ever do more than simply assert the facts they want their hearers to believe. Most models simply assume that the effect of asserting a proposition P is that the hearer believes P. In fact, a speaker may go to great lengths to convince the hearer of the truth of a proposition. She may first assert it, then support it, and even provide support for that support. In such cases, the speaker presumably believes that the combination of utterances is what leads the hearer to accept the main proposition, and we need to model this behavior. Recent work by Galliers (1990), Carberry and Lambert (1999), and Lascarides and Asher (1999) began to address this issue.

Fourth, more integrated accounts of the relationship between the various types of discourse structure are needed. For example, a more detailed understanding of how speakers' intentions are realized via informational structure support more principled and effective text and speech generation. Moreover, a more detailed understanding of the relationship between discourse structure at the segment level and the information structure of Vallduvi (1990) and Steedman (1991) is crucial.

Fifth, over the last decade, statistical techniques have greatly improved performance on tasks such as parsing and word sense disambiguation, and probabilistic models are now widely used for a range of language processing tasks (Jurafsky & Martin, 2000; Manning & Schütze, 1999). We are just beginning to see progress in statistical methods for discourse and dialogue, and there is much work to be done to incorporate more sophisticated models of the discourse phenomena that we wish to approximate with statistical methods.

Finally, with the advent of virtual and lifelike animated agents, many new discourse processing tasks are emerging. Our theories need to be broadened to include spoken language, where much discourse information is conveyed by intonation, and to account for the information contained in a speaker's gestures, which can be used to disambiguate or enrich the meaning of the speaker's utterances (Cassell, Sullivan, Prevost, & Churchill, 2000).

REFERENCES

Aone, C., & Bennett, S. (1995). Evaluating automated and manual acquisition of anaphora resolution strategies. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics* (ACL-95; pp. 122–129). Cambridge, MA: Association for Computational Linguistics.

Appelt, D. E. (1985). *Planning English sentences*. Cambridge, UK: Cambridge University Press.

Asher, N., & Lascarides, A. (1995). Lexical disambiguation in a discourse context. *Journal of Semantics, 12*(1), 69–108.

Asher, N., & Lascarides, A. (1998). The semantics and pragmatics of presupposition. *Journal of Semantics, 15*(3), 239–300.

Asher, N., & Lascarides, A. (2003). *The logic of conversation*. Cambridge: Cambridge University Press.

Asher, N. (1993). *Reference to abstract objects in discourse*. Dordrecht, the Netherlands: Kluwer.

Asher, N., Aurnague, M., Bras, M., & Vieu, L. (1995). Spatial, temporal and spatio-temporal locating adverbials in discourse. In P. Amsili, M. Borillo, & L. Vieu (Eds.), *Time, space, movement: Proceedings of the 5th international workshop*. Château de Bonas, France Groupe Langue, Raisonnement, Calcul, Toulouse.

Barzilay, R., & Elhadad, M. (1997). Using lexical chains for text summarization. In *Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization* (pp. 10–17). Association for Computational Linguistics.

Bateman, J. A., & Paris, C. L. (1989). Phrasing a text in terms the user can understand. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence* (pp. 1511–1517). Detroit, MI.

Breck, E., Burger, J., Ferro, L., House, D., Light, M., & Mani, I. (1999). A Sys called Qanda. In *The Eighth Text REtrieval Conference* (*TREC-8*; pp. 500–246). NIST Special Publication.

Carberry, S. (1983). Tracking user goals in an information-seeking environment. In *Proceedings of the Third National Conference on Artificial Intelligence* (pp. 59–63). Washington, DC.

Carberry, S., & Lambert, L. (1999). A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics, 25*(1), 1–53.

Carletta, J. C. (1992). *Risk taking and recovery in task-oriented dialogue*. Unpublished doctoral dissertation, Edinburgh University.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents*. Cambridge, MA: MIT Press.

Cawsey, A. (1993). *Explanation and interaction: The computer generation of explanatory dialogues*. Cambridge, MA: MIT Press.

Choi, F., Wiemer-Hastings, P., & Moore, J. D. (2001). Latent semantic analysis for text segmentation. In L. Lee & D. Harman (Eds.), *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing* (pp. 109–117).

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris.

Cohen, P. R. (1978). *On knowing what to say: Planning speech acts*. Unpublished doctoral dissertation, Department of Computer Science, University of Toronto. Published by University of Toronto as Technical Report No. 118.

Cohen, P. R., & Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science, 3*, 177–212.

Cohen, R. (1984). A computational theory of the function of clue words in argument understanding. In *Proceedings of the Tenth International Conference on Computational Linguistics* (pp. 251–258). Stanford, CA.

Connelly, D., Burger, J., & Day, D. (1994). A machine learning approach to anaphoric reference. In *Proceedings of the International Conference on New Methods in Language Processing (NEMLAP)*.