Discourse Structure
and the Proper Treatment of Interruptions

Barbara J. Grosz
AI Center
SRI International
Menlo Park, CA
&c CSLI, Stanford University
Stanford, CA

Candace L Sidner
BBN Laboratories
Cambridge, MA

Abstract

This paper reports on the development of a computational theory of discourse The theory is based on the thesis that discourse structure is a composite of three structures the structure of the sequence of utterances, the structure of intentions conveyed, and the attentional state. The distinction among these components is essential to provide adequate explanations of such discourse phenomena as clue words, referring expressions and interruptions. We illustrate the use of the theory for four types of interruptions and discuss aspects of interruptions previously overlooked

1. Introduction

This paper reports on the development of a computational theory of discourse structure that simplifies and extends previous work As we develop it. the theory will be seen to be intimately connected with two nonlinguistic notions, namely intention and attention. Attention and intention are crucial to accounting for the processing of utterances in discourse. Intentions will be seen to play a primary role not only in providing a basis for explaining discourse structure, but also in defining discourse coherence, and providing a coherent notion of the term "discourse" itself

The theory is a further development and integration of two lines of research work on focusing in discourse [6], [7], [B], and more recent work on intention recognition in discourse [[2], [20], [22], [23]] Our goal has been to generalize properly to a wide-range of discourse types the notions of focusing and task structure shown by Grosz to be necessary for processing task-oriented dialogue. One of the main generalizations of previous work will be to show that discourses generally are in some sense "task-oriented," but the kinds of "tasks" that can be achieved are quite varied — some are physical, others mental, others linguistic. As a result, the term "task" is unfortunate, and we will use the more general terminology of intentions — speaking for example of discourse purposes — for most of what we say

Our mam thesis is that the structure of any discourse is a composite of three distinct but interacting constituents: the structure of the actual sequence of utterances m the discourse, a structure of intentions, and an attentional state. The distinction among these constituents is essential to providing an explanation of interruptions (see Section 3), as well as the use of certain types of referring expressions and of various expressions that affect discourse segmentation and structure (discussed in [10]) Most related work on discourse structure (including Reichman [17], Linde [12], Lmde and Goguen [II], and Cohen [4]) conflates at. least two of these constituents. As a result, significant generalizations are lost, and the computational mechanisms proposed are more complex than needed By carefully distinguishing the constituents, we are able to account for the significant observations in this related work while simplifying the explanations given and computational mechanisms used. Related work by Polanyi and Scha ([16], [14], [15]) concentrates on a single component, the linguistic one. and examines in more detail various aspects of its internal structure

In addition to its use in explaining these linguistic phenomena, the theory provides an overall framework m which to answer questions about the relevance of various segments of discourse to each other, and to the overall purposes of the discourse participants Various properties of the intentional component have implications generally for work in natural-language processing. In particular, the range of intentions that underlie discourse is such that approaches to discourse coherence based on selecting discourse relationships from a fixed set of alternative rhetorical patterns are unlikely to suffice in general Furthermore, this study makes evident several problems that must be confronted in extending speech-act related theories (eg, [I], [3], [2], etc.) from coverage of individual utterances to coverage of extended sequences of utterances in discourse

Although a definition of "discourse" must await the development of the theory laid out in the remainder of this paper, some properties of the phenomena we want to explain must be specified now. In particular, we take a discourse to be a piece of language behavior that typically involves multiple utterances and multiple participants. The discourse may be produced by one or more speakers (or writers) and the audience may comprise one or more hearers (or readers) Each *conversational participant* brings to the discourse a set of beliefs, goals, intentions, and other mental attitudes. These attitudes affect a conversational participant's participation in the discourse, they influence both how utterances are produced and how they are understood Where necessary, we use *initialing conversational participant* (ICP) and *other conversational participant* (OCP) to distinguish participants.

2. The Basic Theory

Discourse structure is a composite of three interacting components a linguistic structure, an intentional structure, and an attentional state. These three components of discourse structure deal with different aspects of the utterances in a discourse. Utterances — the actual saying or writing of particular sequences of phrases and clauses — are the basic elements in the linguistic structure. Note that this use of linguistic structure to refer to the structure of a sequence of utterances rather than to single sentence syntactic structure Intentions of a particular sort, namely those whose recognition (by the OCP) is intended (by the 1CP) and which provide the basic reason for the discourse are the basic elements of the intentional structure Attentional state contains information about the objects, properties, relations, and discourse-intentions that are most salient at any given point in a discourse, it summarizes information from previous utterances crucial for processing subsequent ones so that a complete history need not be kept

Together the three constituents of discourse structure provide the information needed by the conversational participants to determine how an individual utterance fits with the rest of the discourse - in essence to figure out why it was said, and what it means, in the context in which it was uttered The context provided by these constituents also forms the basis for certain expectations about what is to come, these expectations too play a role in fitting in new utterances The attentional state serves an additional role, namely it provides the means for actually using the information in the other two structures in the generation and interpretation of individual utterances

2.1. Linguistic Structure

The first component of discourse structure is the structure of the sequence of utterances that form a discourse Just as the words in a single sentence form constituent phrases, the utterances in a discourse are naturally aggregated into *discourse segments* The utterances in a segment, like the words in a phrase, serve particular roles with respect to that segment In addition, the discourse segments, like the phrases, fulfill certain functions with respect to the overall discourse Although two neighboring utterances may be in the same discourse segment, it is also possible for them to be in different segments Likewise two utterances that are not in linear sequence may be in the same segment

The factoring of discourses into discourse segments has been observed across a wide range of discourse types Grosz [6] showed this for task-oriented dialogues Linde [12] found it held for descriptions of apartments, Linde and Gcguen [I I] describe such structuring in the Watergate transcripts Reichman [17] observed it in informal debates, explanations, and therapeutic discourse Cohen [4] found similar structures in essays in rhetoric texts,

There is a two-way interaction between the discourse segment structure and the utterances constituting the discourse linguistic expressions affect the discourse structure, they are also constrained by it, Not surprisingly, linguistic expressions are among the primary indicators of discourse segment boundaries Explicit use of certain words and phrases (e.g.. "in the first place"), and more subtle clues like changes in tense and aspect are among the repertoire of linguistic devices that function wholly or in part to indicate these boundaries ([4], [16], [17]). These linguistic devices can be divided according to whether they indicate changes in the intentional structure or the attentional state of the discourse (or both) The differential use of these linguistic markers provides one piece of evidence for the separation of these two components of discourse structure, In addition, because these linguistic devices function explicitly as indicators of discourse structure, it becomes clear that they are best seen as providing information at the discourse, and not the sentence, level and hence that certain kinds of questions (eg, about their truth conditions) do not make sense

Just as linguistic devices affect structure, so does the discourse segmentation affect the interpretation of linguistic expressions in a discourse Referring expressions provide the primary example of this effect The segmentation of discourse constrains the use of referring expressions by delineating certain points at which there is a significant change in what entities are being discussed. In particular, pronouns and reduced definite noun phrases act differently within a segment than they do across segment boundaries While discourse segmentation is not the only factor governing the use of referring expressions, it is important for capturing one of the constraints on their use Section 2.3 contains some simple examples of the effects of segmentation on referring expressions, more detail can be found in [10]

2.2. Intentional Structure

A rather straightforward property of discourses, namely that they—or, more accurately, those who participate in them- have an overall purpose, turns out to play a fundamental role in the theory of discourse structure In particular, some of the purposes that underlie discourses, and the discourse segments they comprise, provide the means of individuating discourses and of distinguishing coherent discourses from incoherent ones

Although typically the participants in a discourse may have more than one aim in participating in the discourse (eg, a story may entertain its listeners as well as describe an event, an argument may establish someone's brilliance as well as convince that some claim is true), we distinguish one of these purposes as primary to the discourse We will refer to this particular purpose as the discourse *purpose,* or *DP* Intuitively, this discourse purpose is the reason for engaging in this particular discourse * For each of the discourse segments, we can also single out one intention, the *discourse segment purpose,* or *DSP* Intuitively, the DSP says how this segment contributes to achieving the overall discourse purpose **

Typically, an ICP will have a number of different kinds of intentions that lead to initiating a discourse One kind of intention might include intentions to rpeak in a particular language or to utter particular words Another might include intentions to amuse, or to impress The kinds of intentions that can serve as discourse purposes or discourse segment purposes are distinguished from other intentions because they are intended to be recognized (c.f [I], [23]), whereas other intentions are private, that is. the recognition of the DP (or DSP) is essential to its achieving its (intended) effect Discourse purposes and discourse segment purposes share this property with certain utterance level intentions that Grice [5] uses in defining utterance meaning

That is, both why a discourse—a linguistic oct—ond not some other behavior, and why the particular content of this discourse. and not some other information, is being conveyed.

We will assume here a single DP for discourses and DSP for segments. The consequences for the theory of loosening this assumption are discussed in [10].

It is important to distinguish this property from that of being the main intention behind a discourse, a property which the discourse purpose may well not have. Some other intention might be the primary reason for the uttering of a sequence of utterances For example, when on-stage a comedian's main intention may be to amuse He might do this in a variety of ways Some of these could require linguistic behavior — e.g., relate an event sequence, describe a funny object In all of these cases the discourse purpose is the main intention that is *intended to be recognized* (e.g., the intention that the hearers' beliefs come to include some particular beliefs about the sequence of events — those told in the relating •-and their relationship to one another) whereas the intention to amuse is private and need not be recognized by the audience in order for the discourse to succeed

The range of intentions that can serve as discourse, or discourse segment, purposes is open-ended (c.f [25], para 23), much like the range of intentions that underlie purposeful action more generally There is no finite list of discourse purposes, as there is of, say, syntactic categories Thus a theory of discourse structure cannot depend on choosing the DP and DSPs from a small fixed list (as in [17], [19] or [13]), nor on the particulars of individual intentions The particulars of individual intentions are, of course, crucial to understanding any particular discourse, but this is a different issue What is essential for discourse structure is that such intentions bear certain kinds of structural relationships to one another Since the conversational participants can never know the whole set of intentions that might serve as DPs and DSPs, what they must determine are the relevant structural relationships among intentions.

Two structural relationships play an important role. *dominance* and *satisfaction precedence* An action that satisfies one intention, say DSP1, may (be intended to) provide part of the satisfaction of another, say DSP2 When this is the case, we will say that DSPl *contributes to* DSP2, conversely, we will say that DSP2 *dominates* DSPl. For some discourses, including task-oriented ones, the order in which the DSPs are satisfied may be intended to be recognized. DSPl *satisfaction precedes* DSPS in the dominance hierarchy whenever its intention must be satisfied before the other

The following are some examples of the types of intentions that could serve as DPs or DSPs, followed by one particular instance of each type

1. intend that some agent intend to do some physical task, intend that Ruth intend to fix the flat tire

2 intend that some agent (come to) believe some fact, intend that Ruth believe the campfire is started.

3 intend that some agent believe one fact provides support for another, intend that Ruth believe the smell of smoke supports that the campfire is started.

4. intend that some agent intend to identify an object (existing physical object, imaginary object, plan, event, event sequence), intend that Ruth intend to identify my bicycle

5 intend that some agent know some property of an object, intend that Ruth know that my bicycle has a flat tire

DPs and DSPs are basically the same sorts of intentions Whether an intention is a DP or a DSP depends on whether it is the reason for initiating the discourse (in which case it is a DP) or its satisfaction contributes in some way to achieving this main discourse purpose (in which case it is a DSP). Any of the intentions on the preceding list could be either a DP or a DSP. Furthermore, particular instances of any one of them could contribute to another, or to a different instance of the same type. For example, the intention that someone identify some object might dominate several intentions that that person know some property of that object, likewise, the intention to get someone to believe some fact might dominate a number of contributing intentions that that person believe other facts

2.3. Attentional State

The third component of discourse structure, the attentional state, is an abstraction of the focus of attention of the discourse participants as the discourse unfolds. It is inherently dynamic, recording the changes in what objects, properties, and relations are salient at each point in the discourse. The attentional state can be modeled by a set of *focus spaces,* changes m the attentional state are modelled by transition rules for adding and deleting spaces The collection of focus spaces available at any one time we call the *focusing structure,* and the process of manipulating spaces is called *focusing* A focus space is associated with each discourse segment, this space collects together representations of those entities that are salient either because they have been mentioned explicitly m the segment or because they became salient in the process of producing/comprehending the utterances in the segment The focus space also includes the discourse segment purpose, the inclusion of the purpose reflects the fact that the conversational participants are focused on not only what they are talking about but also why they are talking about it

Figure 2-1 illustrates how the focusing structure serves to coordinate the linguistic and intentional structures, as well as capturing the attentional state. The discourse segments (on the left of the figure) are tied to focus spaces (m the middle of the figure) The focusing structure is a stack We illustrate that stack in Figure 2-1 with a pointer between individual focus spaces Information in each space is accessible to other spaces higher in the stack unless otherwise notated with a hash line

The stacking of the focus spaces shown reflects the relative salience of the entities in each space during the corresponding segment's portion of the discourse The stack relationships arise from the ways in which the various DSPs relate, information captured in the hierarchy of DSPs (depicted on the right in the figure) The depiction of spaces shown in the figure is a static representation of what results from a sequence of operations such as pushes onto and pops from a stack A push occurs when the DSP for a new segment contributes to the DSP for the immediately preceding segment. When the DSP contributes to some intention higher in the DSP hierarchy, some number of focus spaces are "popped" from the stack before inserting the new one

Part one of figure 2-1 shows the state of focusing when the paragraph P2 is being processed. Paragraph P1 gave rise to FS1 and had as its discourse purpose DP_1 The properties, objects, relations and purpose

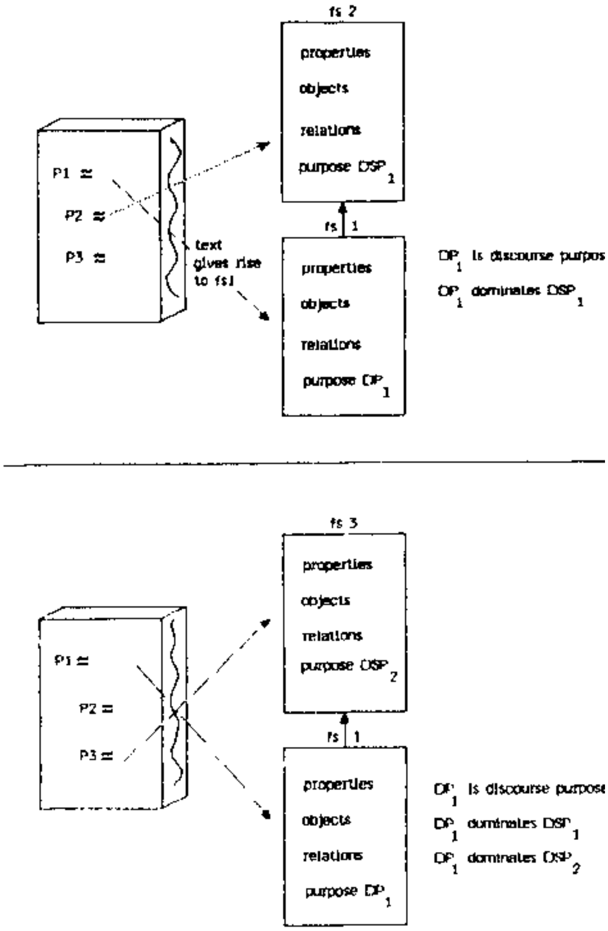

Figure 2-1: Discourse segments, focus spaces and purpose hierarchy

represented in FS1 are accessible but less salient than those in FS2. P2 yields a focus space that is stacked relative to FSI because DP., in FS1 dominates P2's DSP, DSP_1 As a result of the relationship between FSI and FS2, reduced noun phrases will be interpreted differently in P2 than in P_l For example, if some red balls exist in the world and are represented in both FS2 and FSI, "the red ball" used in P2 will be understood to mean that red ball that is represented in *FS2.* If, however, there is a green truck and it is represented only in FSI, "the green truck" occurring in P2 will be understood as that green truck.

Part two of figure 2- 1 shows the state of focusing when paragraph P3 is processed. Because the DSP of FS3, DSP_2, is dominated only by DP_1, and not by DSP_1 FS2 has been popped from the stack, and FS3 has been pushed on

Two essential properties of the focusing structure are now clear First, the focusing structure is parasitic on the intentional structure The relationship among DSPs determines pushes and pops Note however, that which operation is relevant may sometimes be indicated in the language itself For example, the clue word "first" indicates the start of a segment whose DSP contributes to the DSP of the previous segment. Second, the focusing structure, like the intentional and linguistic structures, evolves as the discourse proceeds None of them exists a priori***

The discussion here should also clarify some misinterpretations of focus-space diagrams and task structure in [6], [8] The focus-space hierarchies in that work are best seen as special cases of the attentional state, and the task structure as a special case of the intentional structure we stipulate in this paper Several researchers (eg, Linde and Goguen [II], Reichman [17]) misinterpreted the original research in an unfortunate (and unintended) way -they took the focus-space hierarchy to include (or be identical with) the task structure The conflation of these two structures forces a single structure to contain information about attentional state and intentional relationships It prevents a theory from adequately accounting for certain aspects of discourse including interruptions (see Section 3).

A second confusion was that the task-structure is necessarily a prebuilt tree Taking the task-structure to be a special case of intentional structure makes it clear that the tree structure is simply a more constrained structure than one might require for other discourses, the nature of the task that generates a task-oriented discourse has both dominance and ordering relations,**** while other discourses may not have significant ordering constraints among the DSPs Furthermore there has never been reason to assume that the task structures in task-oriented dialogues are pre-built Rather the task of discourse theory is to explain how the hearer builds up a task structure using information conveyed in the discourse

Figure 2-1 illustrates some fundamental distinctions between the intentional and attentional aspects of discourse structure First, the DP hierarchy provides, among other things, a complete record of the discourse-level intentions and their dominance (and, where relevant, precedence) relations, whereas the focusing structure at any one time can contain only information relevant to a single branch of the hierarchy Second, at the conclusion of a discourse, if the discourse completes normally, the focus stack will be empty while the DP hierarchy will be fully constructed Third, when the discourse is being processed, only the attentional state can directly constrain the interpretation of referring expressions.

It is possible to confuse the DSP with the notion of

Although there ore some rare cases in which one conversational participant has a complete plon for the whole discourse prior to uttering o single word, much more typically, the DSP hierarchy is constructed as the conversational participants create the discourse and need not exist prior to it. It may be more obvious this is true for speakers and hearers of spoken discourse than for readers and writers of texts, but in fact even for the writer, the DSP hierarchy is often developed as the text is wr i 11en.

Even in the task case the orderings may be partial. In fact, the systems built for task-oriented dialogues ([18], [24]) did not use a prebuilt tree, but constructed the tree—based on a partially-ordered model—only as o particular discourse evolved.

center [9] The DSP and center differ in two ways First, the center is an element only of the attentional state, whereas the DSP plays a role in both the attentional and intentional structures. Second, the center may shift within a discourse segment (it almost always shifts across segment boundaries), the DSP does not a change in DSP is what underlies a segment boundary. Although in some cases the intention that is the DSP may be the object that is the center, more typically these do not coincide

In short, the focusing structure is the central repository for the contextual information needed to process utterances at each point in the discourse. It contains those objects, properties, and relations most salient at that point — distinguishing the center from others — and also contains links to those parts of the linguistic structure and the intentional structure that are relevant The ability to identify relevant discourse segments, the entities they make salient, and their DSPs becomes especially important as the amount of information grows over the course of a discourse.

3. Application of the Theory: Interruptions

Interruptions in discourses provide an important test of any theory of discourse structure Because processing an utterance requires figuring out how it fits with previous discourse, it is crucial to figure out which parts of the previous discourse are relevant to it, and which cannot be Thus, the treatment of interruptions has implications for the treatment of the normal flow of discourse Interruptions may take many forms — some are not at all relevant to the main flow of the discourse, others are quite relevant, and many fall somewhere inbetween these extremes A theory must differentiate these cases and explain (among other things) what connections there are between the main discourse and the interruption and how the relationship between them affects the processing of the utterances in both.

The importance of distinguishing between intentional structure and attentional state is evident in the first three examples we consider in this section The distinction also permits us to explain a type of behavior considered by others to be similar — so-called semantic returns--an issue we consider at the end of the section.

The three examples that follow do not exhaust the types of interruptions that can occur in discourse There are additional ways to vary the explicit linguistic and nonlinguistic indicators used to indicate boundaries, the relations among DSPs, and the combinations of focus-space relationships present. These examples illustrate interruptions that fall at different points on the spectrum of relevancy to the main discourse. They can be explained more adequately by the theory of discourse structure given here than by previous theories, and hence provide evidence for the necessity of the distinctions we have drawn.

3.1. Type 1; True Interruptions

The first kind of interruption is the true interruption, a discourse segment whose purpose is distinct from the purpose of the discourse in which it is embedded In the example below, from [15], there are two (separate) discourses, DI indicated in normal type, and D2 in italics.

John came by
and left the groceries
Stop that
you kids
and I put them away
after he left

These two discourses have distinct purposes and convey different information about properties, objects, and relations. Since D2 is embedded within DI, one expects the discourse structures for the two segments to be somehow embedded as well, The theory described in this paper differs from Polanyi and Scha's [14] (and other more radically different proposals as well, e.g., [17]. [4], [11]) in that the embedding occurs *only* in the attentional structure: the focus space for D2 is pushed onto the stack, above (i.e., as more salient than) the focus space for DI, until D2 is completed, as shown in Figure 3—1. The intentional structures for the two segments are distinct There are two DP/DSP structures for the utterances in this sequence It is not necessary to relate these two — end indeed intuitively they are not related

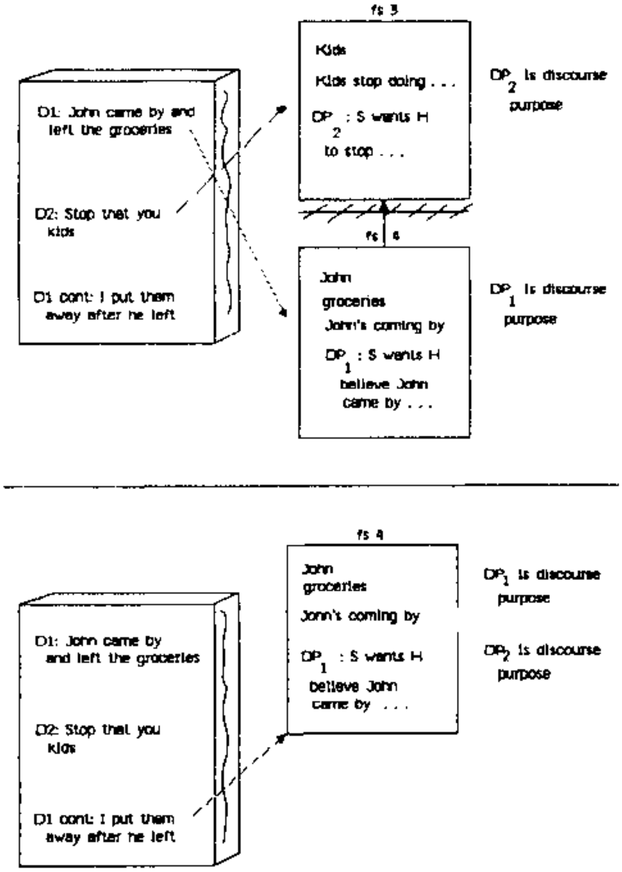
fs 3
Kids
Kids stop doing . . .
DP_2 : S wants H to stop . . .
DP_2 is discourse purpose
D1: John came by and left the groceries
D2: Stop that you kids
D1 cont: I put them away after he left
fs 4
John
groceries
John's coming by
DP_1 : S wants H believe John came by . . .
DP_1 is discourse purpose
fs 4
John
groceries
John's coming by
DP_1 : S wants H believe John came by . . .
DP_1 is discourse purpose
DP_2 is discourse purpose
D1: John came by and left the groceries
D2: Stop that you kids
D1 cont: I put them away after he left

Figure 3-1: The structures of a true interruption.

The focusing structure for true interruptions is different from that for the normal embedding of segments, in that the focusing boundary between the discourse in progress and the interruption is non-penetrable (depicted with a hashed line between focus spaces) The boundary between the focus spaces prevents entities in the one from being available to the other Because the second discourse totally shifts attention to a new purpose (and may shift who the intended hearers are), the speaker cannot use referential expressions in it that depend on the accessibility of entities from the first discourse Because the boundary between the focus space for Dl and that for D2 is non-penetrable, if D2 were to include an utterance like, "put them away", the word "them" would have to refer to something deictically present, and could not be used to refer anaphorically to the groceries.

As the discourse stands however, Dl is resumed almost immediately. The word "them" in "and I put them away" cannot refer to the kids,***** but onlv to the groceries. The focus space for D2 has been popped from the stack Note for this to be clear to the hearer, the speaker must indicate a return to Dl explicitly Two indicators of the "stop that" interruption are assumed to have been present at the time of the discourse -a change of intonation and a change of eye gaze The linguistic indicators are the change of mood to an indicative, and the use of the vocative [16]

Unlike previous accounts, the theory is not forced to integrate these two discourses in terms of a single grammatical structure, nor must the theory provide answers to questions about the specific relationship between segments D2 and Dl, as in [l4] Instend, the intuition readers have of an embedding in the discourse structure is captured in the attentional state by the stacking of focus spaces, which thus accounts for the manner in which the utterances are processed Further, what is intuitively distinct about the two segments is captured in their different intentional (DP/DSP) structures

3.2. Type 2. Flashbacks and Filling in missing pieces

Sometimes a speaker interrupts his or her- own flow of discussion because some purposes, propositions or entities need to be brought into the discourse but have not been the speaker forgot to include those entities first, and now must go back and fill in the missing information. A flashback or- a filler segment results at that point in the discourse These segments contain additional DSPs that must be satisfied before the current DSP can be This type of interruption differs from true interruptions in several ways the DSP for the flashback or filler bears some relationship to the DP for the whole discourse, even though it may not have a close relationship to the DSP of the current segment or to any of the DSPs dominating the current DSP, the linguistic indicator of the flashback or filler typically includes a comment about something going wrong, and the audience always remains the same

In the example below, from [21], the speaker is instructing a mock-up system, played by a person, about how to define and display some knowledge representation information Again, the interruption is indicated by italics

OK Now how do 1 say that Bill is
Woops / forgot about ABC
1 need art indivdual concept for the company ABC

[rcmaiitder of discourse segment on ABC]

Now back to Bill. How do 1 say that Bill is an employee of ABC?

The DP for the whole larger discourse from which this sequence was taken is to provide information about various companies (including ABC) and their employees The outer segment in this example — D-*Bill* — has a DSP--*DSP- Bill*— to tell about Bill, while the inner segment- --*D-ABC*- -has a DSP ■■ *DSP-ABC--\.o* convey certain information about ABC. Because of the nature of the information being told, there is order to the final structure of the DP DSPs information about ABC must be conveyed before all of the information about Bill can be The speaker in this instance does not realize this constraint until after he begins The "flashback" interruption allows him to satisfy DSP ABC while suspending satisfaction of DSP Bill (which he then resumes) Hence, as shown in Figure 3-1.', there is an intentional structure rooted at DP and with DSP-ABC and DSP-Bill as ordered sister- nodes

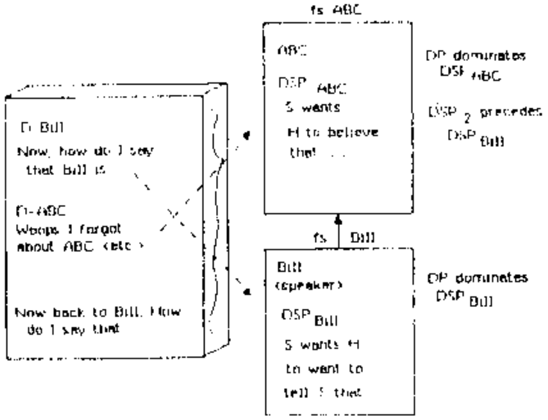

Figure 3-2: The structures of a flashback.

The available linguistic data permit two possible attentional states as appropriate models for flashback type interruptions The simpler model has a focusing structure identical to the one that would ensue if the flashback segment were a normally embedded segment, as depicted in Figure 3-2 The focus space for the flashback- - *FS-ABC**—is pushed onto the stack above the focus space for- the outer segment——*FS-Bill,* and all of the entities in both focus spaces are normally accessible for reference The more complex model uses an auxiliary stack FS -Bill (and possibly some additional spaces) are put onto the auxiliary stack for the duration of the interruption After an explicit indication that there is a return to work on DSP-Bill (e.g.. the "Now back to Bill" used in this example), any focus spaces left on the stack from the flashback are popped off, and all spaces on the auxiliary stack (including F S - Bill) are returned to the main stack

The major difference between these two models is that the first allows entities relevant to the interrupted material to be accessible during the interruption whereas in the second thev are not Which model is correct depends on whether in the embedded segment (D-ABC) the speaker can refer to Bill or other entities in FS-Bill using less than full definite descriptions

Because this is so clearly the case on other grounds, the segment boundary is clear even to a reader after the fact.

Because the use of pronouns seems to be connected much more with centering than with focus space boundaries, the appropriateness of pronominal reference to Bill ("he") is not an adequate test, as a result the current example — and other data available — do not indicate a clear choice between these models However, the explicit return to D —Bill in this example suggests the more complex model is needed

This kind of interruption is distinct from true interruptions because there is a connection, although indirect, between the DSPs for the two segments. Further the linguistic markers of the start of the interruption indicate that there is a precedence relation between these DSPs (and hence the need for the correction). Flashbacks are also distinct from normally embedded discourses by the precedence relationship between the DSPs for the two segments, and the order in which the segments occur. The second attentional model further distinguishes flashbacks from normal discourse because it provides for information being saved but not accessible (in the auxiliary stack) during the interruption

3.3. Type 3: Digressions

The third type of interruption we consider, which we call a digression, is a segment that is linked to the segment it interrupts by some entity that is salient in both, but that has a DSP unrelated to the DP to which the interrupted segment's DSP contributes For example, if while discussing Bill's role in company ABC, one conversational participant interrupts with, "Speaking of Bill, that reminds me, he came to dinner last week," Bill remains salient, but the DSP changes The salient object on which the digression is based might be the DSP, but more typically is some object, relation, or property in the focus space for the interrupted segment A typical means of beginning such digressions are phrases like "speaking of John" and "that reminds me '

In processing digressions, the DSP for the digression forms the base of a separate intentional structure just as in the case of true interruptions A new focus space is formed and pushed onto the stack, but it contains at least one — and possibly other — entities from the interrupted segment's focus space. Like the flashback-type interruption, the digression usually must be closed with an explicit closing utterance such as "getting back to ABC . "

3.4. Noninterruptions — "semantic returns"

One case of discourse behavior which we must distinguish are the so-called "semantic returns" discussed by Polanyi and Scha [16] In all the interruptions we have considered there is a need to pop the stack when the interruption is over, and the mam flow of the discourse is resumed. The focus space for the interrupted segment is "returned to.' In the semantic, return case, entities and DSPs previously salient are taken up once again, but they are explicitly reintroduced The state of the focus stack is not a factor in constraining such "returns " For example, suppose yesterday two people had discussed how badly Jack behaved at the party, and then today one says "Remember our discussion about Jack at the party? Well, a lot of other people thought he acted just as badly as we thought he did." The utterances today call up, or return to, yesterday's conversation through the intention that more be said about Jack's poor behavior, but the return is not a return to a previous focus space

Anything that can be talked about once, can be talked about again later. However, if there is no focus space on the stack corresponding to the segment and DSP being discussed further, then, as Polanyi and Scha [16] point out. there is no popping of the stack The separation of attentional state and intentional structure makes clear what is occurring in such cases, and the intuitions that lie behind the use of the term "semantic return." In re-introducing some entities from a previous discourse, conversational participants are establishing some connection between the DSP of the new segment and the intentional structure of the original discourse It it not a return to a previous focus space because the focus space is gone from the stack and the items to be referred to must be explicitly re-established. It is a return, at least in some sense, to a previous intentional structure.

4. Conclusions and Future Research

The theory of discourse structure presented in this paper generalizes from theories of task-oriented dialogues. It differs from previous generalizations in carefully distinguishing three components of discourse structure—one linguistic, one intentional, and one attentional The distinctions are crucial for an explanation of interruptions, clue words, and referring expressions

The particular intentional structure used also differs from the analogous aspect of previous generalizations. Although, like them it provides the backbone for the discourse segmentation and determines structural relationships for the focusing structure (part of the attentional state), unlike them it does not depend on the particular details of any single domain or discourse — type.

Although (obviously) not complete, the theory provides a solid basis for investigating not only discourse structure, but also discourse meaning, and for constructing discourse-processing systems. Several difficult research problems remain to be addressed Of these, we take the following two to be of primary importance

1 What is the relationship between discourse-level (DP/DSP)$_0$ and utterance-level (speech acts) intentions^0

2. What information do discourse participants use to recognize these intentions, and how do they do it?

Finally, the theory suggests two important conjectures First, that a discourse is coherent only when its discourse purpose is shared among the conversational participants, and when each of the utterances of the discourse contributes to achieving this purpose, either directly or indirectly by contributing to the satisfaction of a discourse segment purpose. Second, that the notion of "topic" is primarily an intentional notion, it is best seen as referring to the DP, and DSPs. Previous discussions of the "topic" of an utterance or discourse have been confused because uses of the term "topic" have variously referred to notions that are essentially syntactic (e.g., the "wa" marking in Japanese, surface subject in English), attentional (the center of an utterance), and intentional (the DSP of a segment)

Acknowledgments

The research reported in this paper has been made possible by a gift from the System Development Foundation, and by a grant from the Defense Advanced Research Project Agency under contract N00014-77-C-0378

References

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

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

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

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

[5] Grice, HP
Utterer's Meaning and Intentions
Philosophical Review 68(2): 1 47-1 77, 1969

[6] Grosz, B.J.
Discourse Analysis
In D. Walker (editor), *Under standing Spoken Language,* chapter IX, pages 235-268 Elsevier North-Holland, New York City, 1978

[7] Grosz, B.J
Focusing in Dialog
In *Theoretical Issues in Natural Language Processing- 2,* pages 96-103 The Association for Computational Linguistics, University of Illinois at Urbana Champaign, July. 1978.

[8] Grosz. B.J
Focusing and Description in Natural Language Dialogues
In A. Joshi, B Webber. I Sag (editors), *Elements of Discourse Understanding,* pages 84-105.Cambridge University Press, 1981

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

[10] Grosz, B.J. and Sidner, C.L
The Structures of Discourse Structure
In *Discourse Structure.*Ablex Publishers, 1986

[11] Linde, C and Goguen, J
Structure of Planning Discourse
J. Social Biol. Struct. 1 219-251, 1978

[12] Linde, C.
Focus of Attention and the Choice of Pronouns in Discourse
In T. Givon (editor). *Syntax and Semantics. Vol. 12 of Discourse and Syntax,* pages 337-354.Academic Press. Inc., 1979.

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

[14] Polanyi, L. and Scha. R.
A Syntactic Approach to Discourse Semantics
In *Proceedings of Int'l Conference on Computational Linguistics* Stanford University, Stanford, CA, 1984

[15] Polanyi, L and Scha, R.
A Model of Natural Language Discourse.
In *Discourse Structure.Able** Publishers, 1986

[16] Polanyi, L, and Scha, R. J. H.
On the Recursive Structure of Discourse.
In *Proceedings of the January 1982 Symposium on Connectedness in Sentence, Text, and Discourse* The Catholic University of Tilsburg, Tilsburg, The Netherlands, 1983
In press

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

[18] Robinson, A.
Interpreting verb phrase references in dialogs
In *Proceedings of the Third Biennial Conference of the Canadian Society for Computational Studies of Intelligence* Victoria. May, 1980.

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

[20] Sidner, C.L, and Israel, D.J
Recognizing intended meaning and speaker's plans.
In *Proceedings of the International Joint Conference in Artificial Intelligence,* pages 203-208 IJCA1, IJCAI. Vancouver, B.C . August, 1981

[21] Sidner, C.L
Protocols of Users Manipulating Visually Presented Information with Natural Language
Technical Report 5128, Bolt Beranek and Newman Inc., September, 1982

[22] Sidner, C.L
What the Speaker Means The Recognition of Speakers' Plans in Discourse
Computers and Mathematics with Applications 9(1), 1983.
Special Issue on Computational Linguistics - Nick Cercone, guest editor

[23] Sidner, CL
Plan parsing for intended response recognition in discourse.
Computational Intelligence 1(1) 1—10, February, 1985

[24] Walker. D
Understanding Spoken Language.
Elsevier North-Holland, New York City, 1978.

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