

**The Structure of  
Multiple-Headed Negotiations**

**Carolyn Penstein-Rosé**

June 1995

Report CMU-LCL-95-1

**Laboratory for  
Computational  
Linguistics**

139 Baker Hall  
Department of Philosophy  
Carnegie Mellon University  
Pittsburgh, PA 15213-3890

University Libraries  
Carnegie Mellon University  
Pittsburgh, PA 15213-3890

# The Structure of Multiple Headed Negotiations

Carolyn Penstein Rosé

Computational Linguistics Program

Carnegie Mellon University

cprose@clcl.cmu.edu

## Abstract

In this paper I demonstrate that theories which postulate a strict tree structure of discourse on either the intentional or attentional level are not totally adequate for handling spontaneous dialogues. I use as evidence analyses of the structure of multiple headed negotiations. These are negotiation dialogues in which multiple suggestions are being negotiated at once. The question I explore in this paper is how to characterize the structure of a discourse in which multiple non-embedded adjacency pairs are initiated within a single turn since it is unclear in these cases how response parts are matched up with their corresponding initiating parts if discourse is tree structured. I argue that each non-embedded initiating part exists on a separate thread of the discourse so that an initiating part can be matched with its corresponding response part on its own thread. I develop my theory for the structure of these dialogues in the spirit of (GS86) and (Loc94). I support my claim that discourse is structured as a tapestry of interwoven threads by demonstrating that discourse level parallel structure makes it necessary to model attentional state as a graph structured stack rather than as a simple stack in order to make correct predictions about which discourse entities are readily available for pronominal reference, particularly with deictic pronouns like "that" and "those". I demonstrate that the graph structured stack model is a generalization of the simple stack model and therefore covers all of the phenomena covered by the simple stack model in addition to the data discussed in this paper which is problematic for the simple stack model. An implementation of this theory is discussed in (RELED95).

## 1. INTRODUCTION

Conversation Analysts have traditionally characterized the structure of spoken discourse, in particular the relationship between speech acts, in terms of adjacency pairs (SS73; SSJ74; Gof76; Cou77) where first pair parts are matched with an expected second pair part. For example, questions are paired with their answers, and greetings are often paired with similar greetings, and suggestions are paired with acceptances or rejections. In this paper, I analyze the structure of multiple headed negotiations. These are negotiation dialogues in which multiple suggestions are being negotiated at once. The question I explore in this paper is how to characterize the structure of a discourse in which multiple non-embedded first pair parts are presented within a single turn<sup>1</sup> since it is unclear in that case how second pair parts are matched up with their corresponding first pair parts if discourse is tree structured. I will demonstrate that theories which postulate a strict tree structure of discourse on either the intentional or attentional level are not totally adequate for handling spontaneous dialogues.

The corpus which my model is based on is composed of 50 recorded and transcribed dialogues in which speakers attempt to schedule a meeting together. Although both speakers share the intention of scheduling a meeting together, the speakers differ with respect to which days and which times are preferable for them to meet according to their respective schedules. So in these dialogues, speakers discuss their potential intentions to meet at particular times on particular days. It is these potential intentions which they do not necessarily share.

The question is how the structure of the turn in which multiple topics are presented affects the possibilities for the structure of subsequent turns which address the same topics. In particular, I will explore how this structure constrains possibilities for discourse pronominal reference. The role which discourse structure plays in constraining resolution of pronominal reference has been explored in (Pol88; Fox87; Ash93; GS85). I will develop my theory of discourse structure in the spirit of (GS86; Loc94) which has played an influential role in the analysis of discourse entity saliency and in the development of dialogue processing systems.

The examples below, all of which were extracted from my corpus with the exception of (IV) and (V) which were constructed for the purpose of illustrating the difference between presenting multiple suggestions

---

<sup>1</sup>A turn is defined as the largest span of continuous talk by a single speaker.

in a single clause versus presenting them each in separate clauses, demonstrate how speakers present multiple suggestions in parallel within a single turn. Each parallel suggestion expresses a potential intention to meet at a particular time.

In example (I), s1 suggests two different Tuesdays in parallel and s2 rejects both of them. In order for the reference to "those" in unit (4) to be resolved, both suggested meeting times must be available for pronominal reference.

(I)

s1: (1) What about Tuesday afternoons?

(2) The twenty third I'm free from three to five.

(3) And on the thirtieth I'm free after noon.

s2: (4) Okay, both of those are bad for me.

In example (II), s1 again suggests two meeting times, the fourth and the sixth. In order for the reference to "that" to be resolved in unit (3), the meeting time on the sixth must be relatively more salient than the meeting time on the fourth. On the other hand, we know that both meeting times must be salient since the reference to "those" in (6) can be resolved. The reference to "that" in unit (5) is ambiguous. It could potentially refer either to the meeting on Friday or both the meeting times.

(II)

s1: (1) Okay, so the fourth is a potential.

(2) Or how about Friday the sixth in the morning?

(3) Do you have any time free that day?

s2: (4) No.

(5) that won't work.

(6) Both of those days are bad.

In example (III), s1 suggests that next week is a good time to meet and then proceeds to make two

parallel suggestions which make this suggestion more specific. Then s2 rejects the first of these suggestions and accepts the other.

(III)

- s1: (1) Next week is pretty good for me.  
(2) I think next Wednesday afternoon would be best,  
(3) or Thursday after two.
- s2: (4) Wednesday doesn't seem to be too good.  
(5) Thursday is.  
(6) Anywhere from two thirty to five on Thursday is fine.

In example (IV), s1 makes three parallel suggestions, each in different units. In unit 4, in order to resolve the reference to "that", the suggested meeting time on Wednesday must be relatively more salient than the other two.

(IV)

- s1: (1) I am free Monday.  
(2) Tuesday is good as well.  
(3) I could also make it Wednesday.
- s2: (4) That day won't work  
(5) because I have class on Wednesday.  
(6) But either Monday or Tuesday would work.

In contrast to example (IV), the unnatural example (V) indicates that when the suggested times are presented in a list in the same unit, it isn't the case as it is when they are presented in separate units that the last suggested time is relatively more salient than the other two.

(V) ??

s1: (1) I am free Monday, Tuesday, and Wednesday.

s2: (2) That day won't work

(3) because I have class on Wednesday.

(4) But either Monday or Tuesday would work.

In the following sections, I will use this data to argue that a simple stack model of attentional state, where it accounts for many interesting discourse phenomena, is not adequate for handling the case where there is discourse level parallelism as in these examples. I will argue that a graph structured stack model makes more correct predictions, handling all of the examples above.

### 1.1. The Grosz and Sidner Account

Grosz and Sidner's model is composed of three separate but interacting components: the linguistic structure, the intentional structure, and the attentional state. The linguistic structure is the structure of the sequence of sentences which make up the discourse. It is on this level that the discourse is divided into segments. Each discourse segment is associated with a discourse segment purpose. The relationships between these discourse segment purposes make up the intentional level. There are two relationships discussed in their 1986 paper, namely *dominates* and *satisfaction-precedes*. A discourse segment purpose *dominates* a second one when satisfaction of the second one contributes to the satisfaction of the first one. A discourse segment purpose *satisfaction-precedes* another discourse segment purpose if the first purpose must be satisfied before the second one can be attempted to be satisfied. The attentional state is modeled with a stack where each discourse segment is associated with a focus space<sup>2</sup> which is pushed on the stack when that discourse segment is processed. The attentional stack determines which entities can be referred to pronominally, generally those entities on or near the top of the focus stack.

For each discourse segment, a focus space is pushed onto a focus stack with all of the entities which are evoked in that segment. The dominance relations in the intentional structure determine which operations are performed on the attentional stack when a discourse segment is processed. If discourse segment purpose

---

<sup>2</sup>The focus space contains all of the referents evoked within the corresponding discourse segment.

A is dominated by discourse segment purpose B then focus space A can be pushed on top of focus space B if focus space B is on top of the stack when focus space A is pushed. Otherwise, if discourse segment purpose A is not dominated by discourse segment purpose B, focus space B must be popped from the stack when focus space A is pushed. When a focus space is pushed onto the attentional stack, it is pushed on top of the focus space associated with the discourse segment purpose its corresponding discourse segment purpose contributes to. All of the focus spaces in between the new one and the one which it will be pushed on top of are popped off prior to pushing the new focus space. Pronominal references within a discourse segment are computed after the segment's focus space has been pushed on to the stack.

In a recent extension to Grosz and Sidner's original theory, described in (Loc94), each discourse segment purpose represents an intention that the speakers form a shared plan (GK93). These discourse segment purposes are expressed in terms of the intention operator *Int. That*, one of four intention operators discussed in (GK93) based on the theory of intentional behavior and practical reasoning discussed in (BIP88). *Int. That* represents the intention that some proposition hold. Another intention operator discussed in (GK93) is *Pot. Int. That*. This represents an agent's potential intention that some proposition hold. Potential intentions are used to account for an agent's process of weighing different means for accomplishing an action he is committed to performing (BIP88). These potential intentions are not discourse segment purposes in Lochbaum's theory since they cannot form the basis for a share plan having not been decided upon yet. It is not until they have been decided upon that they become *Int. That's* which can then become discourse segment purposes.

For example, deliberation over how to accomplish a shared plan can be represented as an expression of multiple *Pot. Int. That's*, each corresponding to different alternatives. For each factor distinguishing these alternatives, the potential intentions are all discussed inside of a single discourse segment whose purpose is to explore the options so that the decision can be made. Lochbaum discusses an example of a dialogue containing examples of such deliberation in (Loc94), included here in Figure 1. In each of the three subdialogues, a different type of information is discussed. In the first one, two alternative ways of opening a savings account are introduced. In the second one, the interest rates of each are compared. In the third, required initial deposits of each are discussed. Both alternatives are discussed in each subdialogue.

- (1) Customer: I'd like to open a savings account.
- (2) What types do you offer?
- (3) Teller: Passbook and investment.
- (4) Customer: What's the interest rate on your passbook account?
- (5) Teller: 2.5 %.
- (6) Customer: And the rate for the investment account?
- (7) Teller: 3.0 %
- (8) Customer: Okay.
- (9) How big are the initial deposits for the two accounts?
- (10) Teller: \$1000 for the passbook and \$5000 for the investment.
- (11) Customer: Okay.
- (12) Whom do I see to open a passbook account?

Figure 1: Example of Deliberating Over Selecting a Plan for Action

The stipulation that *Int. That's* can be discourse segment purposes but *Pot. Int. That's* cannot has a major impact on the analysis of scheduling dialogues such as the one in Figure 2 since the majority of the exchanges in scheduling dialogues are devoted to deliberating over which date and at which time to schedule a meeting. In our corpus, the only factor which distinguishes good meeting times from bad ones is whether the speakers are busy doing something else at those times or not. This would seem to leave all of the deliberation over meeting times within a single monolithic discourse segment. This leaves the vast majority of the dialogue with no segmentation. This would leave us with the question of how to account for shifts in focus which seem to occur within the deliberation segment as evidenced by the types of pronominal references which occur within it. For example, in the dialogue presented in Figure 2, how would it be possible to account for the differences in interpretation of "Monday" and "Tuesday" in (3) with "Monday" and "Tuesday" in (14)? It cannot simply be a matter of immediate focus since the week is never mentioned in (13). And there are no semantic clues in the sentences themselves to let the hearer know which week is intended. Either there is some sort of structure in this segment more fine grained than would be obtained if *Pot. Int. That's* cannot be discourse segment purposes, or another mechanism must be proposed to account for the shift in focus which occurs within the single segment. I argue that rather than propose an additional mechanism, it is more perspicuous to lift the restriction that *Pot. Int. That's* cannot be discourse segment purposes. In my theory a separate discourse segment is allocated for every potential plan discussed in the dialogue.



- (1) S1: We need to set up a schedule for the meeting.
- (2) How does your schedule look for next week?
- (3) S2: Well, Monday and Tuesday both mornings are good.
- (4) Wednesday afternoon is good also.
- (5) S1: It looks like it will have to be Thursday then.
- (6) Or Friday would also possibly work.
- (7) Do you have time between twelve and two on Thursday?
- (8) Or do you think sometime Friday afternoon you could meet?
- (9) S2: No.
- (10) Thursday I have a class.
- (11) And Friday is really tight for me.
- (12) How is the next week?
- (13) If all else fails there is always video conferencing.
- (14) S1: Monday, Tuesday, and Wednesday I am out of town.
- (15) But Thursday and Friday are both good.
- (16) How about Thursday at twelve?
- (17) S2: Sounds good.
- (18) See you then.

Figure 2: Example of Deliberating Over A Meeting Time

It could potentially be argued that it does not make sense to model discourse segment purposes as potential intentions because a speaker would not say something unless that speaker was committed to it somehow and that it is necessary to use intentions instead. However, I argue that this is not a problem with our representation and that it in fact makes more sense to model the discourse segment purposes in question as potential intentions.

It is clear that in making a suggestion, a speaker is at least committed to making that suggestion. However, in Lochbaum's theory, what the speaker is committed to in engaging the other speaker in a discourse segment is something different. She writes that "The purpose of the segment is taken to be the intention that the discourse participants form that plan." (Loc94) (p. 49). Because according to (BIP88) a rational agent will not be committed to two things which are inconsistent, a rational agent will not intend to form two plans which are mutually exclusive. Since a single meeting has only one meeting time, if a speaker intends to form a plan to meet at a particular time, it does not make sense that that speaker would also

intend to form a plan to meet at a different time unless these two intentions corresponded to two separate meetings. It would make sense, however, for an agent to potentially intend to form two different alternative plans for a meeting.

It might be argued that the process of negotiating over alternative meeting times is somehow analogous to the types of knowledge precondition dialogues discussed in (Loc94). This process is most similar to her subdialogues in which the purpose is to achieve a satisfactory description of a plan. This satisfactory description would contain in its completion a time and date on which to meet in the case of scheduling dialogues. It cannot be the case that the speaker intends that the time and date in question be suggested since simply suggesting a meeting time and date does not directly contribute to building a plan for a meeting. So the speaker must have some sort of intention about the outcome of suggesting a particular time and date. However, it cannot be the case that for each potential meeting time and date the speaker intends that the corresponding time and date be acceptable instantiations in order to achieve a satisfactory description of a plan for a meeting since this would not allow the speaker to engage in subdialogues about mutually exclusive alternatives in parallel since that would make it necessary for the speaker to have inconsistent intentions. One might suggest as the only alternative left that the purpose is to come to some agreement on the status of the time and date as a possible acceptable description for the meeting plan. But intending to establish some unspecified status of a suggestion has the same problems as intending to make a suggestion.

Formulating each parallel negotiation segment as having the purpose of a potential intention that the discourse participants form a plan to meet on the specified time and date does not have these problems. Since it is acceptable for a speaker to have mutually exclusive potential intentions, this avoids the problem of the speaker intending that the suggested meeting time and date be acceptable. And since the role of potential intentions in the deliberation process is already built in to Grosz and Kraus's SharedPlans framework, it would avoid the problem of the speaker simply intending to suggest a particular time and date. The question which remains is whether it makes sense for a potential intention being sufficient motivation for a speaker to actually say something. I argue that it is.

In (GK93), it is theorized that an agent weighs a number of potential intentions internally in the process of deliberating over multiple alternatives before making a commitment. In order to weight these alternatives,

an agent must have some basis by which to evaluate them. In some cases, an agent does not have enough information to make the decision. One such case is when the speaker has part of the information required in order to determine if a particular instantiation is appropriate and the other speaker has the other part of the information. In my theory of the negotiation process in these circumstances, the agent can invite the other agent to participate as an equal party in her internal deliberation process by making it explicit by verbalizing it and allowing the other agent to respond. This is the process which I am modeling when I say that discourse segment purposes within these types of negotiation segments are potential intentions. Thinking of the negotiation process this way makes using potential intentions as discourse segment purposes make sense since it gives the speaker a motivation for verbalizing something which she is not specifically committed to.

## 1.2. An Alternative Account

A number of different potential intentions are discussed in Figure 2. In (2), the potential intention to meet sometime next week is expressed. In (3), the more specific potential intention to meet either Monday or Tuesday in the morning is expressed. In (4), the potential intention to meet on Wednesday is expressed. Each time a speaker expresses an attitude towards a potential plan, a potential intention is expressed. Other potential intentions are expressed throughout this dialogue.

By allocating a separate discourse segment to each potential plan, a situation arises where discourse segments seem to overlap one another in some cases. For example, if the discussion about the potential meeting on Thursday is in a separate segment from the potential meeting on Friday, then either the two segments are interleaved such that (5), (7), and (10) are in one segment and (6), (8), and (11) are in another, or all six of these sentences are in separate segments altogether. Which of these options are selected would make definite predictions about which entities are available for pronominal reference at different points between sentence (5) and sentence (11).

A similar issue is addressed in (Web88) where the following example is discussed:

House A is in Palo Alto, House B in Portola Valley. Both were built in 1950, and both have three bedrooms. House A has two baths, and B, 4. House B also has a kidney shaped pool. House

A is on a quarter acre, with a lovely garden, while house B is on 4 acres of steep wooded slope, with a view of the mountains. The owner of house A is asking \$425K. The owner of house B is asking \$600K. That's all I know about house A. This I heard from a friend, who saw house B before it came on the market. (p 115)

In this example, because the reference of "that" in the second to last sentence seems unnatural to resolve to the description of house A and the interpretation of "this" as referring to the description of house B seems very strange, Webber concludes that in discourses where there is discourse-level parallel structure as in the example above, the individual parallel components cease to be available for pronominal reference<sup>3</sup>. I argue that it is not the parallel structure which makes these references difficult to resolve but rather a combination of two factors. The first factor is that because the description of house B was most recently modified, it is more in focus than the description of house A at the end of the two descriptions. So if a pronominal reference is made to one or the other description directly after the description of B is modified, it will most naturally refer to the description of B. Secondly, when "this" is used along side of "that" it is communicated that what is referred to with "this" is closer to the speaker than what is referred to with "that". So it would seem very odd to use "this" to shift the focus. Because the focus was shifted to the description of house A in the previous sentence, it seems odd to use "this" to shift the focus back to the description of house B. If this is the reason why those two references are unnatural, I would predict that reversing the last two sentences would make the two references seem more natural and would indicate that it is not the case that the individual descriptions are not available for pronominal reference after the two descriptions are given in parallel. And this turns out to be the case.

House A is in Palo Alto, House B in Portola Valley. Both were built in 1950, and both have three bedrooms. House A has two baths, and B, 4. House B also has a kidney shaped pool. House A is on a quarter acre, with a lovely garden, while house B is on 4 acres of steep wooded slope, with a view of the mountains. The owner of house A is asking \$425K. The owner of house B is

---

<sup>3</sup>Some readers have indicated that "this" seems to refer to the whole description including both house A and house B. This is an alternative reading. The point I am addressing here, however, is simply the ability of "this" in this context to refer to the description of house B.

asking \$600K. This I heard from a friend, who saw house B before it came on the market. And that's all I know about house A.

This indicates that each parallel description is maintained throughout the description, supporting the view that it is two discourse segments interleaved<sup>4</sup> rather than several fragmentary discourse segments in a sequence. I will argue based on a corpus analysis of 50 dialogues that the model in which the discourse segments are interleaved makes more correct predictions about possibilities for discourse pronominal reference than the model where a totally separate discourse segment is created for each non-contiguous portion of the discourse.

This issue raises interesting questions about the representation of attentional state. Discourse segment purposes neither dominate nor are dominated by parallel discourse segment purposes by definition. Therefore, according to Grosz and Sidner's theory, focus spaces associated with one parallel thread must all be popped from the top of the stack when focus spaces associated with another parallel thread are pushed on. This predicts that pronominal references can never be made which include entities from focus spaces associated with more than one parallel thread at the same time. I will demonstrate that this prediction is false.

---

<sup>4</sup>i.e. I am saying that the description of house A is in a separate segment from the description of house B.

## 2. METHODOLOGY

In this section I will discuss my methodology for analyzing the dialogues used as data for the theory developed in this paper. I will explain how I derive the representation for the attentional state from this representation. In the next section I will use this coding scheme to demonstrate that incorrect predictions are made if attentional state is modeled as a simple stack structure and that correct predictions are made if it is modeled as a graph structured stack.

I first introduced this methodology in (Ros95). The coding scheme presented here is an extension of the coding scheme presented in that paper. My goal for this coding scheme is that it should be reliable in the sense that the analyses which I produce with this coding scheme should be reproducible. My coding scheme is based on (Edm81) in which Edmondson develops a formalism for analyzing adversarial conversations in scenarios such as "Speaker1 wants Speaker2 to do X" or "Speaker1 did X and X is bad for Speaker2". What distinguishes Edmondson's formalism from those of others who have studied sequences of speech acts and how speech acts relate to one another in discourse is that he has separated the notion of illocutions from that of interactional structure which others have left conflated (Sea92; Kla73; LF77; SS73; SC75). In his formulation, the illocution is an action which communicates the beliefs, desires, and intentions communicated by an utterance without reference to what function this action plays in the discourse. The interactional component represents the relationship between this action and the surrounding discourse, how this sentence works together with surrounding sentences in order to produce conversational outcomes, i.e. what is accomplished through a unit of discourse.

What I am coding with my coding scheme is similar to the interactional structure in Edmondson's work. From this structure, I derive my equivalent of Grosz and Sidner's intentional structure. My analyses with this coding scheme identify moves in dialogues and their relationships with one another. With this coding scheme I have carved out a finite set of discrete relationships which I identify in the data I analyze. I have not made any assumptions about how these relationships can be ordered or how many relationships a single sentence<sup>5</sup> can participate in.

---

<sup>5</sup>The smallest unit of analysis in my coding scheme is actually a unit called a pseudo sentence unit which is generally a clause. More details about pseudo-sentence segmentation are given below.

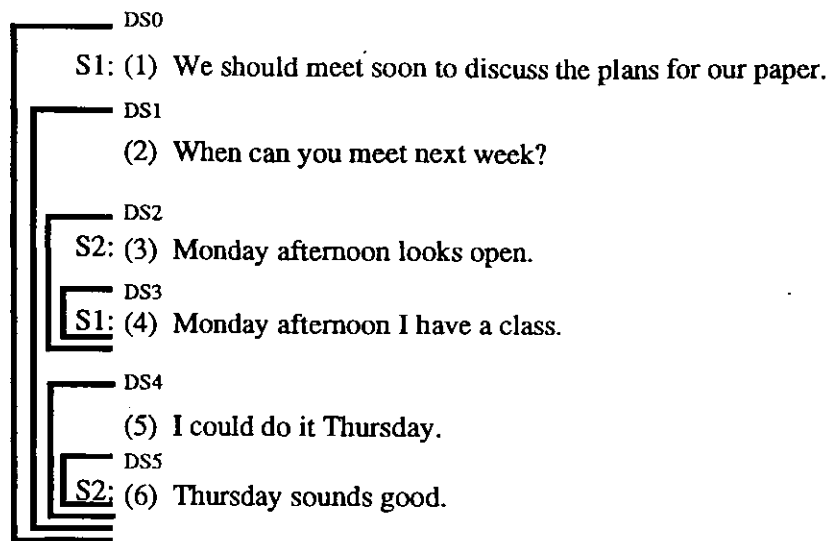


Figure 3: Example Scheduling Dialogue

## 2.1. The Coding Scheme

I will explain my coding scheme by demonstrating how to apply it to the example dialogue in Figure 3. In this figure I have noted how I would segment this discourse considering that potential intentions can act as discourse segment purposes. At the end of this section I will describe how to derive this segmentation from the coding scheme.

The discourse segment purpose of DS0 is a shared intention for the two speakers to settle on a date on which to hold their meeting. The discourse segment purpose of DS1 is S1's potential intention to make that meeting during the next week. The period of time denoted by 'next week' is only a fraction of the period of time in which the meeting could be held according to sentence (1). So this potential intention is dominated by the more general intention of finding a time for scheduling a meeting. Likewise, the potential intention to meet on Monday is more specific than meeting next week, so this requires another subordinate discourse segment. When S1 responds to S2's suggestion, a new subordinate discourse segment must again be allocated because an intention to respond to a potential intention contributes to determining the status of that potential intention (i.e. whether it is shared by the other speaker or not). S1 then makes a suggestion to meet on Thursday. This potential intention is not subordinate to the potential intention of meeting on Monday. It is dominated by the potential intention to meet next week. And because it is more specific to

meet on Thursday than to meet next week, a new subordinate discourse segment is allocated inside of DS1. Finally, sentence (6) is placed in DS5 which is subordinate to DS4 since sentences (5) and (6) were expressed by different speakers. This is simply an intuitive justification for this segmentation. After the coding scheme is introduced, concrete rules for deriving the segmentation from the coding scheme will be described.

My coding scheme was inspired by Edmondson's (81) dissertation research discussed above. In Edmondson's coding scheme, utterances were related to one another through conversational moves. Through these moves, the utterances worked together to produce conversational outcomes. These conversational outcomes were said to be produced through the interactional structure, but they were never formalized, nor was the process of deriving them from the interactional structure. In my coding scheme, I have defined what these outcomes are and how they are derived in order to make it possible to determine whether any given analysis for a dialogue is correct for a given interpretation. Also, rather than assign each sentence a label indicating how it fits into the interactional structure as Edmondson did, I assign these labels to relationships between conversational outcomes. There are two types of conversational outcomes, local outcomes and non-local outcomes. Local outcomes are produced by single pseudo-sentence units. And non-local outcomes are produced as these local outcomes interact with previous outcomes. The relationships between outcomes are conversational moves. The set of relationships between outcomes makes up the interactional structure. These relationships make it possible for a sentence to play multiple roles in the interactional structure.

**OUTCOME:** 3-b  
**SPEAKER:** speaker1  
**START TIME:** next week, Monday afternoon  
**DURATION:** two hours  
**TOPICS:** plan for the paper  
**ATTITUDE:** positive

Figure 4: Sample Outcome

Because scheduling dialogues such as the one in Figure 3 are what I am applying my coding scheme to, I consider conversational outcomes to be partial or full plans for meetings which are either positive, indicating that the speaker or speakers are in favor of this plan, or negative, indicating that the speaker or speakers