

Reconciling Initiative and Discourse Structure

Susan E. Strayer and Peter A. Heeman

Department of Computer Science and Engineering
OGI School of Science and Engineering
Oregon Health & Science University
20000 NW Walker Rd Beaverton, OR 97006, USA
susan_strayer@yahoo.com heeman@cse.ogi.edu

Abstract

In this paper we consider how initiative is managed in dialogue. We propose that initiative is subordinate to the intentional hierarchy of discourse structure. In dialogues from the TRAINS corpus we find that inside a segment initiated by one speaker, the other speaker only makes two types of contributions: a special kind of acknowledgment we call *forward acknowledgments*, and short contributions that add content to the segment. The proposal has important implications for dialogue management: a system only needs to model intentional structure, from which initiative follows.

1 Introduction

The dialogue manager of a spoken language system is responsible for determining what contributions a system can make and when it can make them. The question is, what should the dialogue manager pay attention to in order to accomplish this? Two areas of research have shaped our understanding of what happens in dialogue: research in dialogue structure and in mixed initiative.

Grosz and Sidner (1986) proposed a theory of discourse structure to account for why an utterance was said and what was meant by it. Their theory has three components: linguistic structure, intentional structure and attentional state. Intentions are key to accounting for discourse structure, defining discourse coherence, and “providing a coherent conceptu-

alization of the term ‘discourse’ itself.” The *intentional* structure describes the purpose of the discourse as a whole, and the relationship of the purpose of each discourse segment to the main discourse purpose or other discourse segment purposes. All utterances within a segment contribute to the purpose of that segment. This theory, however, does not comment on initiative within the segment. Nor does it specify when and how speakers should start a segment or end the current one. Hence, it underspecifies what speakers can do in dialogue.

Research in initiative works to account for which speaker is driving the conversation at any given point. For example, in a question-answer pair, the speaker asking the question is said to have the initiative (Whittaker and Stenton, 1988; Walker and Whittaker, 1990; Novick and Sutton, 1997). Whittaker and Stenton segmented dialogues where initiative shifts from one speaker to the other. They found that initiative “did not alternate from speaker to speaker on a turn by turn basis, but that there were long sequences of turns in which [initiative] remained with one speaker. In a mixed initiative system, the dialogue manager needs to track initiative in order to know when the system should add significant content, and when it should let the user take over. However, no theory has offered a good account of why a speaker would want to take the initiative, or keep it once they have it.

In the rest of this paper we first describe previous work in discourse structure and in initiative and describe our coding of them. Next, we explore the relationship between discourse structure and initiative. As previous

studies have found (Whittaker and Stenton, 1988; Walker and Whittaker, 1990), there is a close correlation between them, but the relationship is not direct. We then explore how initiative can shift within a subdialogue and find two types of contributions that a speaker can make in a discourse segment: a special kind of acknowledgment we call *forward acknowledgment*, and short contributions that add content to the segment. We propose that initiative is subordinate to intentional structure. Additionally, our proposal is better able to account for question-answer pairs and how initiative returns to the original speaker after an embedded subdialogue. It will have important implications for dialogue management: a system only needs to model intentional structure, from which initiative follows.

2 Discourse Structure and Initiative Analysis

Our proposal for managing initiative builds on two main areas of research, discourse structure and initiative. We start by discussing the work of Grosz and Sidner (1986), which ties speaker’s intentions to linguistic structure, then discuss the work of Whittaker, et al. in initiative. We introduce our coding of three dialogues in the TRAINS corpus, a corpus of human-human task-oriented dialogues, in which two participants work together to formulate a plan involving the manufacture and transportation of goods (Allen et al., 1995; Heeman and Allen, 1995). In these dialogues, one speaker is the user (u), who has a goal to solve, and the other speaker is the system (s), who knows the detailed information involved in how long it takes to ship and manufacture goods.

2.1 Discourse Structure

Discourse structure is used to analyze dialogue from the top down, starting with the purpose of the discourse as a whole, then the purpose of each discourse segment, in order to understand how each utterance fits into the dialogue. The theory of discourse structure developed by Grosz and Sidner (1986) proposes that discourse structure is made up of

three components: linguistic structure, intentional structure, and attentional state. Our work focuses on the first two components. The *linguistic structure* is a hierarchical segmentation of the dialogue into *discourse segments*. Segment boundaries are identified by changes in tense and aspect, pause lengths, speech rate, and discourse markers, such as “anyway,” “by the way,” “so,” and “first of all.” The *intentional structure* is a hierarchy of segment purposes. Each discourse segment has a purpose, and the purpose of each segment contributes to the purpose of its parent. Intentional structure is key to understanding what the discourse is about and explains its coherency.

Subdialogue coding: In our study, the first author segmented dialogues into subdialogues based on the purpose of the utterance (Smith and Gordon, 1997; Traum and Hinkelman, 1992). We established two classes of subdialogues: *task subdialogues*, segments that describe subtasks in the dialogue, and *clarification subdialogues*, local segments that clarify a gap in understanding, either to request missing information or to supply missing information to the other speaker. The segment *initiator* gives the first utterance in the segment and establishes its purpose.¹ The left side of Figure 1 gives an example of a discourse segment (or subdialogue) with two embedded clarification subdialogues, and the segment initiator for each subdialogue. Generally, we found the dialogue structure in the dialogues we analyzed to be quite flat, with few embedded structures. Typically, task subdialogues occurred at the same level, and clarification subdialogues were embedded in a task subdialogue, as seen in the example in Figure 1.

2.2 Initiative

Initiative is held by the speaker who is driving the conversation at any given point in the conversation (Whittaker and Stenton, 1988,

¹The segment initiator corresponds to the initiating conversational participant (ICP) of Grosz and Sidner’s theory. The non-initiator corresponds to the *other conversational participant* (OCP).

Segment	Initiator	Speaker	Utterance	Initiative
	u	u	where pick up um one of the tankers there umph- with uh oranges that will be used	u
		s	mm-hm	
		u	for the t-	u
	s	s	okay to move the oranges we need a boxcar	s
		u	okay	
		u	we'll bring a boxcar from Elmira with the engine three	u
		s	okay	
		u	um so we'll get to Corning with the engine three and the boxcar will get to Corning and we'll pick up uh	u
	u	u	can a tanker and a boxcar be pulled by an engine	u
		s	yes	
		u	so okay so we'll pick up a tank of uh the tanker of OJ orange- oranges	u
		s	ok- right	

Figure 1: Example of Discourse Structure and Initiative Segments (d921-5.2: utt25-utt36).

Walker and Whittaker, 1990; Novick and Stenton, 1997). It has been used to analyze discourse from the bottom up, starting with utterances. We start with adjacency pairs (Schegloff and Sacks, 1973), which consist of a *first part*, uttered by one of the speakers, and a *second part*, uttered by the other. The first part sets up expectations for the second part, and hence the speaker of the first part can be viewed as being in control of the dialogue during both parts of the adjacency pair. Below we give the annotation scheme used by Whittaker, et al. (Whittaker and Stenton, 1988; Walker and Whittaker, 1990) for annotating initiative based on utterance type.

Assertions: Declarative utterances used to state facts. The speaker has initiative, except when it is a response to a question.

Questions: Utterances intended to elicit information from others. The speaker has initiative, except when it follows a question or command.

Commands: Intended to induce actions in others. The speaker has initiative.

Prompts: Utterances with no propositional content (e.g., “yeah,” “okay”). These utterances do not exhibit initiative.

Whittaker and Stenton used the initiative codings as a basis for segmenting dialogues. They used dialogues between an expert and

a client about diagnosing and repairing software faults. They found that not only did initiative pass back and forth between the speakers (unlike single-initiative dialogues), but that initiative often stayed with a speaker for on average of eight speaker turns.

Whittaker and Stenton (1988) looked at the correlation of control boundaries to discourse markers, and Walker and Whittaker (1990) looked at anaphoric reference. These are the same kinds of linguistic evidence that Grosz and Sidner (1986) said marks discourse segment boundaries. In fact, Walker and Whittaker claimed that initiative segments are the discourse segments of Grosz and Sidner’s theory, with the speaker with initiative being the initiator of the segment, who establishes the discourse segment purpose. However, they acknowledged that “there can be topic shifts without change of initiation, change of [initiative] without topic shift.” In fact, when we look at the dialogue excerpt given in Figure 1, we see that the initiative segmentation identified the first subdialogue, but not the second. However, Walker and Whittaker did not specify the relationship between initiative and discourse structure.

Initiative coding In our study, the first author coded initiative using the annotation scheme of Whittaker and Stenton (1988). The right side of Figure 1 shows the annotation for utterances where speakers demonstrate initiative. For utterances where the speaker does

Table 1: Correlation of Segment Boundaries

	Initiative vs Subdialogue	Segment Initiator vs Subdialogue	Initiative vs Segment Initiator
Boundaries	113	113	46
Hits	47	46	41
Misses	66	67	5
False Positives	35	0	41
Recall	42%	41%	89%
Precision	57%	100%	50%

not demonstrate initiative, initiative is said to belong to the last speaker that demonstrated it. Hence, when the system uttered “mm-hm” in the second utterance, which does not demonstrate initiative, initiative does not change from the user. We also show initiative segment boundaries, which are identified by changes in initiative. Here, we see that initiative swings back and forth between the system and the user several times, leading to three initiative segments.

3 Relationship between Initiative and Discourse Structure

Walker and Whittaker (1990) suggested that changes in initiative correspond to changes in discourse structure, but they did not determine the exact relationship between them. In this section we analyze the differences between initiative segments and discourse structure for three dialogues from the TRAINS corpus. We find that there is a close relationship, but not a direct one.

3.1 Segment Boundary Comparison

In this section, we compare initiative boundaries (where initiative shifts from one speaker to the next) to subdialogue boundaries (where a new subdialogue begins) using recall and precision.² An initiative boundary is scored as a hit if there is a corresponding subdialogue boundary. It is scored as a false positive if there is no subdialogue boundary. A subdialogue boundary is scored as a miss if it has no initiative boundary. For example,

²Recall = Hits / (Hits + Misses)
Precision = Hits / (Hits + False Positives)

in Figure 1, there are two hits (the boundary between the 3rd and 4th utterances and the boundary between the 5th and 6th utterances), two misses (the boundary between the 8th and 9th utterances and the boundary between the 10th and 11th utterances), and no false positives. The second column of Table 1, “Initiative versus Subdialogues”, gives the results. We see that both recall and precision are very low for initiative boundaries relative to discourse boundaries. However, comparing initiative segments to discourse segments is not fair. The misses in Figure 1 should be expected since the initiator of the last subdialogue is the same as the higher level subdialogue.

To show the effect of the unfairness, we contrast changes in segment initiator to discourse segment boundaries in the third column of Table 1, “Segment Initiator versus Subdialogue.” Not surprisingly, we obtained a precision of 100%: by definition, the segment initiator is only set at the beginning of each discourse segment. However, we only obtained a recall rate of 41%. This means only 41% of discourse segment boundaries are initiated by a different speaker. We should not expect these boundaries to have a change in initiative, since there is no change in segment initiator. A fair comparison should contrast changes in initiative only to changes in discourse segment initiator. The results of doing this is shown in the fourth column, “Initiative versus Segment Initiator.” Here we see much better results for recall; however, precision is still very low. We will return to the low precision rates in Section 4.

Table 2: Initiative Held by Segment Initiator

	Total	Clarification Subdialogues	Task Subdialogues
Subdialogues	91	45	46
First utterance	91 (100%)	45 (100%)	46 (100%)
Whole subdialogue	77 (85%)	45 (100%)	32 (70%)

Segment			
Initiator	Initiative	Speaker	Utterance
u	u	u	okay so we have to take oranges from Corning and bring them to Elmira
		s	right
	u	u	and then back to Bath by
	s	s	+ by noon +
	u	u	+ mid- + by noon

Figure 2: Forward acknowledgment (d92a-5.2: utt14-18).

3.2 Shifts Within Discourse Segments

Although the recall rate in the last column of Table 1 is very good, it shows there are some changes in discourse segment initiator that are not matched by changes in initiative. The question is, does this happen at the beginning of the segment, the end, or in the middle? We looked at how often the segment initiator is the same as the speaker with initiative for the first utterance in each discourse segment. As seen in Table 2, this does give us a 100% correct rate, meaning that a discourse segment can only be initiated by the speaker also taking the initiative. This is not unexpected, since the speaker needs to contribute something new, otherwise it would not count as the beginning of a new discourse segment. However, the initiative does not always stay with the initiator for the entire segment, as seen in the last row of Table 2.

4 Reconciling Initiative Inside Discourse Segments

The first utterance of each discourse segment shows perfect agreement between the initiator of the segment and speaker with initiative, as seen in Table 2. But what happens during the course of the segment? In this section we focus on subdialogues where the non-initiator

makes a contribution and the initiator finishes the segment.

4.1 Forward Acknowledgments

In the TRAINS corpus there are times when listeners were so synchronized with the speaker that could they anticipate what the speaker was going to say and fill it in before the he said it. Figure 2 gives an example, where the system filled in “by noon” for the user. Typical acknowledgments are utterances that indicate understanding of an utterance made by the other speaker and do not contribute content. Our phenomena of *forward acknowledgments* also indicate understanding, but of what the other speaker is about to say, even before he says it. By filling in what the other speaker was about to say, the speaker indicates understanding and also moves the conversation forward. In both examples of forward acknowledgments in the three dialogues, the initiator finished the utterance of the other speaker.

Forward acknowledgments are coded as demonstrating initiative, because they add content. However, this initiative is subordinate to the initiative of the main segment, so they are show as being embedded in the parent segment.

Segment	Initiator	Initiative	Speaker	Utterance
	u	u	u	and then take those uh t- to Dan- and then go to Dansville
		s	s	so that's + uh let's see and + that's one one more hour
			u	+ and +
		u	u	yeah and we can un- we can
		s	s	drop off at + the +
		u	u	+ drop + off th- that boxcar and take well
			u	dr- yeah drop off the boxcar + of +
		s	s	+ and then + take two empty ones
	u		u	right two empty ones down to Avon
			s	+ oh +
	u		u	+ and + pick up the the bananas
			s	right

Figure 3: Other contribution (d93-19.5, utt83-utt93)

4.2 Other-Contributions

The rest of the utterances coded with initiative made by the non-initiator of the segment were more substantial contributions. Here, the speaker added content to the discourse segment that is not predicted from the initiator's speech. We refer to these as *other-contributions*, and they often occur where the two speakers are closely collaborating and are highly synchronized.³ In Figure 3, we show a dialogue excerpt in which the two speakers are so closely synchronized that they pick up parts of each others utterances and build on it. Initiative shifts back and forth between the two speakers, but, in fact, we think this phenomenon of other-contributions is related to the phenomena that Schiffrin (1987) referred to as *shared turns*.

4.3 Effect on Initiative

Table 3 shows what happens to initiative after the non-initiator makes a contribution demonstrating initiative. There were 25 of

³It is interesting to note the amount of overlapping speech (marked with a '+') in these examples of other-contributions. It might be the case that when speakers are highly synchronized, they are more bound to loosen the restrictions on turn-taking. This is something we hope to investigate in the future.

Table 3: After Contribution by Non-Initiator

Contributions by Non-Initiator	25
New subdialogue	7
Embedded subdialogue	1
Initiative returns to segment initiator	14
Initiative held by non-initiator	3

these cases. In 7 of them, the next utterance demonstrating initiative occurred in a new subdialogue, and in one case it occurred in an embedded subdialogue. We focus on the remaining 17 cases where the next utterance demonstrating initiative occurred within the same discourse segment. In 14 of these, initiative returned to the segment initiator, usually in the very next utterance (13 out of 14). In only 3 cases does initiative stay with the non-initiator. This result is contrary to previous theories of initiative (Walker and Whittaker, 1990; Chu-Carroll and Brown, 1997), which would expect initiative to stay with the non-initiator.

4.4 Discussion

Forward acknowledgments and other-contributions are exceptions to the general rule that initiative tends to reside with the

same speaker. Based on the results of our small preliminary study, we propose that initiative is subordinate to the intentional structure of the discourse theory of Grosz and Sidner. Initiative is held by the segment initiator. The non-initiator can make utterances that contribute to the purpose of the current discourse segment, namely forward acknowledgments and other-contributions, but initiative remains with the segment initiator. Hence, initiative does not need to be tracked, because it is held by the initiator of the discourse segment.

This proposal allows either speaker to contribute to the purpose of a discourse segment, which accounts not only for forward acknowledgments and other-contributions, but also can account for embedded subdialogues and the answer to a question in a question-answer pair. It can be argued that embedded subdialogues initiated by either speaker contribute to the purpose of its parent subdialogue. For example, in Figure 1, the subdialogue initiated by the system (s) contributes to the purpose of the parent subdialogue by pointing out a problem with the user’s plan to move oranges, which the user then adjusts. The clarification subdialogue initiated by the user (u) checks that the plan the user is developing will work, and so it also supports the general purpose of developing the plan.

The proposal also can also be used to simplify how initiative is used in question-answer pairs. Whittaker and Stenton (1988) assigned initiative to the speaker of statements, except when it was the answer to a question, in which case it belonged to the speaker asking the question. In our view, a question-answer pair is a subdialogue with the initiative belonging to the questioner. The answer is coded with initiative, but it is an other-contribution, and so this initiative is subordinate to the questioner’s initiative.

Other researchers have struggled with structure in initiative. Chu-Carroll and Brown (1997) referred to initiative as *dialogue initiative*, and proposed a second level, *task initiative*, to model who is adding domain actions. In contrast to our proposal, which

makes initiative subordinate to intentional structure, they proposed that dialogue initiative is subordinate to their task initiative. Hence, their model could incorrectly predict who has initiative after the non-initiator makes a contribution. As was shown in Table 3, after the non-initiator makes an other-contribution within a subdialogue, generally initiative returns to the segment initiator of the subdialogue, instead of staying with the non-initiator. Chu-Carroll and Brown also used task initiative to model how cooperative a system should be. With novice users, the system would tend to have task initiative and thus make domain actions, but not so with experts. This is similar to Smith and Gordon’s (1997) use of four levels of initiative, which set how much initiative was given to the system and how much was given to the user in one of four levels. Although a system needs to reason about how helpful it needs to be, it is unclear whether this can be done through a single variable that is tied to dialogue initiative.

5 Conclusion

In this paper, we have proposed that initiative is subordinate to intentional structure in dialogue. We have backed up this claim by examining utterances that demonstrate initiative made by the non-initiator of the discourse segment. We found that after these utterances, initiative returns to the segment initiator in almost all cases. The reconciliation of initiative and discourse segments means that we now understand how initiative and dialogue level intentions are related and have a clearer picture of how both participants can contribute to discourse intentions.

Based on our results, initiative in itself does not need to be tracked. Initiative belongs to the speaker who started the current discourse segment. Therefore, a dialogue manager only needs to model intentional structure.

6 Future Work

Our preliminary study was based on a small set of data coded by one of the authors. Plans

for future work include annotating more dialogues with multiple coders to ensure the results reported here are reproducible. We also intend to analyze what happens to initiative after an embedded subdialogue is completed to verify that initiative stays with the segment initiator of the parent segment. We also want to better understand the few cases where stays with the non-initiator (the three cases in Table 3).

In addition to more TRAINS dialogues, we will code dialogues from other corpora, such as Maptask (Anderson et al., 1991) and Switchboard (Godfrey et al., 1992). This will help ensure that we do not introduce idiosyncrasies of the TRAINS corpus into our theory. Rather than just code initiative, we will use the DAMSL annotation scheme (Core and Allen, 1997). This scheme annotates the *forward* and *backward*-looking functions of each utterance, from which initiative can be derived. Reliable intercoder agreement has been obtained with this coding scheme. For coding discourse structure, several schemes have been proposed (Passonneau and Litman, 1997; Flammia, 1998; Nakatani et al., 1995; Traum and Nakatani, 1999) ranging from coding flat segmentation on monologues to hierarchical segmentation of dialogues. We will use these annotation schemes as a foundation, and monitor our annotation results to ensure we achieve good intercoder reliability.

Our theory necessitates that we better understand the structure of discourse, how it is built, and the actions and rules that a discourse manager can use to affect the discourse structure. We also need to understand the reasoning process that determines whether a participant will make an other-contribution or start a new subdialogue. Since dialogue is a collaborative effort (Cohen and Levesque, 1994; Clark and Wilkes-Gibbs, 1986), we also need to explore how the participants collaborate on the discourse structure.

7 Acknowledgments

The authors gratefully acknowledge funding from the Intel Research Council. The authors also thank David Traum and members

of the Centers for Spoken Language Understanding and Human Computer Communication at OGI for helpful discussions and comments.

References

- J. Allen, L. Schubert, G. Ferguson, P. Heeman, C. Hwang, T. Kato, M. Light, N. Martin, B. Miller, M. Poesio, and D. Traum. 1995. The Trains project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7:7–48.
- A. Anderson, M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert. 1991. The HCRC map task corpus. *Language and Speech*, 34(4):351–366.
- J. Chu-Carroll and M. Brown. 1997. Tracking initiative in collaborative dialogue interaction. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Spain, July.
- H. Clark and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.
- P. Cohen and H. Levesque. 1994. Preliminaries to a collaborative model of dialogue. *Speech Communications*, 15(3–4):265–274, December.
- M. Core and J. Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *Working notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*.
- G. Flammia. 1998. Discourse segmentation of spoken dialogue: an empirical approach. Doctoral dissertation, Department of Electrical and Computer Science, Massachusetts Institute of Technology.
- J. Godfrey, E. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the International Conference on Audio, Speech and Signal Processing (ICASSP)*, pages 517–520.
- B. Grosz and C. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- P. Heeman and J. Allen. 1995. The Trains spoken dialog corpus. CD-ROM, Linguistics Data Consortium, April.

- C. Nakatani, B. Grosz, D. Ahn, and J. Hirschberg. 1995. Instructions for annotating discourse. Technical Report 21-95, Center for Research in Computing Technology, Harvard University, Cambridge MA, September.
- D. Novick and S. Sutton. 1997. What is mixed-initiative interaction? In *1997 AAAI Spring Symposium on Computational Models for Mixed Initiative Interaction*, pages 24–26, Stanford University, March. AAAI Press.
- R. Passonneau and D. Litman. 1997. Discourse segmentation by human and automated means. *Computational Linguistics*, 103–139.
- E. Schegloff and H. Sacks. 1973. Opening up closings. *Semiotica*, 7:289–327.
- D. Schiffrin. 1987. *Discourse Markers*. Cambridge University Press, New York.
- R. Smith and S. Gordon. 1997. Effects of variable initiative on linguistic behavior in human-computer spoken natural language dialogue. *Computational Linguistics*, 23(1):141–168.
- D. Traum and E. Hinkelman. 1992. Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599. Special Issue on Non-literal language.
- D. Traum and C. Nakatani. 1999. A two-level approach to coding dialogue for discourse structure: Activities of the 1998 working group on higher-level structures. In *Proceedings of the ACL'99 Workshop Towards Standards and Tools for Discourse Tagging*, pages 101–108, June.
- M. Walker and S. Whittaker. 1990. Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 70–78.
- S. Whittaker and P. Stenton. 1988. Cues and control in expert client dialogues. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 123–130.