DISCOURSE MARKER USE IN TASK-ORIENTED SPOKEN DIALOG*

Donna K. Byron
Department of Computer Science
University of Rochester
Rochester NY 14604, U.S.A.
dbyron@cs.rochester.edu

Peter A. Heeman
France Télécom CNET
Technopole Anticipa - 2 Avenue Pierre Marzin
22301 Lannion Cedex, France.
heeman@lannion.cnet.fr

ABSTRACT

Discourse markers, also known as cue words, are used extensively in human-human task-oriented dialogs to signal the structure of the discourse. Previous work showed their importance in monologues for marking discourse structure, but little attention has been paid to their importance in spoken dialog systems. This paper investigates what discourse markers signal about the upcoming speech, and when they tend to be used in task-oriented dialog. We demonstrate that there is a high correlation between specific discourse markers and specific conversational moves, between discourse marker use and adjacency pairs, and between discourse markers and the speaker's orientation to information presented in the prior turn.

1 INTRODUCTION

In human-human task-oriented dialogs, *discourse markers* abound. Discourse markers are a linguistic devise that speakers use at the beginning of a contribution to signal its relationship to the current discourse state [2, 3, 7, 8]. For instance, discourse markers can be used to mark changes in the global discourse structure, as exemplified by 'by the way' to mark the start of a digression and 'anyway' to mark the return from one. However, in task-oriented dialogs, most discourse marker usage concerns local cohesion. Previous work showed the importance of discourse markers in monologues, but little attention has been paid to their importance in spoken dialog systems. Since discourse markers occur in turn-initial position, the recognition and understanding of these markers can help narrow the alternatives for what will follow.

This paper describes the results of annotating five dialogs from the Trains corpus [4]. The dialogs contained a total of 471 utterances, 401 turns, and 24.5 minutes of speech. For the purposes of this study, only turn-initial speech acts which began with a discourse marker were studied; discourse marker phrases internal to a turn were not considered. We first examine how discourse markers can be used as a timely indicator of how the following speech will relate to the current discourse state. We then examine the circumstances under which speakers typically use discourse markers, first in terms of adjacency pairs, which capture local discourse expectations, then in terms of grounding, which captures how speakers display their continuing understanding of a dialog.

*This research was completed while both authors were at the University of Rochester. The authors would like to thank James Allen. This material is based on work supported by NSF grant IRI-9623665. A more detailed report is available at http://www.cs.rochester.edu/trs/ai-trs.html.

2 CONVERSATIONAL MOVES AND DISCOURSE MARKERS

Previous work on discourse markers proposes that they correlate with discourse moves and are used to signal discourse structure. A speaker's choice of discourse marker in turn-initial position helps the listener develop an expectation of the discourse intent of the upcoming utterance.

In the annotated dialogs, only 'and', 'so', 'well', and 'oh' occurred frequently enough to merit analysis. The discourse function or intent of the utterances has been classified into "conversational moves". Table 1 summarizes the conversational move of turns from the annotated dialogs with discourse markers in turn-initial position. The table shows that each discourse marker is strongly correlated with particular conversational moves.

	Turns beginning with			
Conversational Move	**And**	**Oh**	**So**	**Well**
Restate	0	0	6	0
Summarize Plan	5	0	4	0
Request for summary	1	0	3	0
Conclude	0	0	15	0
Elaborate Plan	22	0	0	0
Correction	0	0	0	7
Respond to new info	0	17	0	0

Table 1: Correlations with conversational move

This section presents examples of each conversational move with representative dialog segments for each discourse marker. In the dialog segments, the 'user' (turns labeled 'u') and the 'system' (turns labeled 's'), both humans, collaborate to plan train routes to deliver specified cargo to specified towns within stated time constraints. These dialog examples demonstrate that turn-initial discourse markers provide strong clues about the relation of the upcoming speech to the current discourse state.

2.1 Conversational moves beginning with 'Oh'

Prior work suggests that a speaker employs 'oh' to mark that he has undergone a change in his current state of knowledge, information, orientation, or awareness [6]. 'Oh' can mark a self-initiated repair, or it can come after clarification, correction, or a response to a question [8].

In the Trains corpus, 'oh' marks a change in its speaker's orientation to information. 'Oh' turns frequently occur after the other speaker provides new information.

utt20	u:	how far is it from Elmira to Bath
utt21	s:	two hours
utt22	u:	**oh** really so then w- we could actually take like Engine E two have it go to Bath...

Ex. d92-1: Oh signals incorporation of new information

When used in this way 'oh' connotes not only that its speaker understands information provided in the previous turn, but also that the information was somehow contrary to his prior beliefs. 'Oh' can also signal that information provided in a prior turn was incomplete or incorrect. In the example below, the 'system' had agreed that three boxcars were in Dansville, but then realizes that there are also boxcars in Elmira. His use of 'oh' in utt53 signals that the rest of the turn describes a change in his informational state.

utt50	u:	um there are three boxcars in Dansville
utt51	s:	yep
utt52	u:	um
utt53	s:	**oh** there's also two in Elmira
utt54	u:	two in Elmira **oh** um hm okay

Ex. d92-1: Oh signals a change of informational state

'Oh' turns may allow a discontinuity in the topic, implying that the reason for changing the topic is because the speaker has experienced a sudden change in his beliefs about some domain-related information.

2.2 Conversational moves beginning with 'Well'

Schiffrin suggests that 'well' is used almost exclusively at the beginning of a response, to signal that an upcoming contribution is not fully consonant with the set of possible responses implied by the question initiator.

utt16	u:	how long would it take to load the oranges from the warehouse into the engine
utt17	s:	uh **well** we can't load oranges into an engine we need a boxcar

Ex. d92-3.1: Canonical use of **well** in a response

In the Trains dialogs, 'well' is typically used to correct a misconception or to suggest an alternative plan. It is found not only at the beginning of responses, but also after the other speaker has just stated a fact or drawn a conclusion with which the current speaker is about to disagree.

utt53	u:	and then I'm done
utt54	s:	**well** you have to get to Avon still

Ex. d93-17.4: Well is used to make corrections

Starting a turn with 'well' sets up an expectation that the speaker is about to disagree with or correct information, either implicit or explicit, from the prior turn.

2.3 Conversational moves beginning with 'And'

'And' is used extensively in turn-initial position in the Trains dialogs. As Schiffrin's analysis predicts, 'and' is used primarily to mark that the current utterance is a continuation of the same speaker's prior turn. 'And' also correlates strongly, though not absolutely, with the presentation of new information. Out of 28 'and'-initial turns in the annotated dialogs, 20 contributed new information.

utt68	u:	fill up the boxcar with the oranges
utt69	s:	okay
utt70	u:	**and** pick up a tanker and bring it back to Elmira
utt71	s:	okay
utt73	u:	**and** make the OJ right
utt74	s:	mm-hm
utt75	u:	**and** then fill up the tanker
utt76	s:	uh-huh
utt77	u:	**and** then go to uh from Elmira to Avon via Corning and Dansville

Ex. d93-17.2: And signals continuation of the prior turn

After part of the plan has been constructed in utt62 - utt68, the 'user' continues adding onto the plan with 'and' turns. It can sometimes be difficult to tell the discourse use of 'and' from its lexical function as a conjunction. Elaborations of the plan typically involve describing a sequence of events, so speakers naturally conjoin them with 'and'. But notice that in the above example the 'user' is delivering information in installments. Each turn is already assumed to be a continuation of the list begun in the prior turn. The turn-initial 'and' can be removed without changing the logical flow or informational content of the turns, so it must be operating at the discourse level.

2.4 Conversational moves beginning with 'So'

'So' is traditionally said to mark main idea units, such as the return to a main level after a clarification sub-dialog or to bring a higher-level context space back into focus [8]. In the Trains corpus, 'so' turns provide conclusions, summaries, or restatements. When a restatement begins with 'so', the listener does not expect the upcoming utterance to build new information onto the emerging plan. 'So' contrasts with 'and' in interesting ways which are discussed in the next section.

The 'system' participants in Trains dialogs tend to take frequent checkpoints to compare the emerging plan with time requirements in the problem statement. These summaries typically begin with 'so'.

utt41	s:	okay **so** it'll get to Dansville at ten a.m.and then to Corning **so** get to Corning at eleven a.m.

Ex. d93-10.4: So marks a conclusion about the plan

'So' is used not only to present conclusions about the plan, but also to request that the other speaker contribute a conclusion about the plan when the current speaker does not have the information to make the conclusion himself.

utt37	u:	hm let me think here there are no boxcars at Avon right
utt39	s:	there're no bo- right
utt40	u:	hm
utt42	s:	**so** what exactly ar- are you trying to do **so** your goal is
utt43	u:	okay well the goal is transport two boxcars...

Ex. d93-9.3: So used to request a summary of the plan

Partners in the Trains dialogs often re-state information that has just been presented while they are thinking about it, and these utterances invariably begin with 'so'.

utt57	u:	and then when it gets to Corning
utt58	s:	yep
utt59	u:	it'll leave one of the boxcars of bananas
utt60	s:	okay **so** we're going to get to Corning and leave a boxcar of bananas

Ex. d93-10.4: So marks a restatement of old information

In the Trains corpus, when a misunderstanding or miscommunication sets up the need for a clarification subdialog or side discussion, return to the main topic is typically marked by 'so'. Another common use of 'so' in Trains is when the speaker presents a summary of the plan to re-establish a prior context space, then adds onto the plan. In that case the re-stated information will begin with 'so', and the new part of the plan will be marked with 'and'.

2.5 Contrasts between 'And' and 'So'

'And' and 'so' are the most frequently occurring discourse markers in the Trains dialogs, so it is interesting to contrast their use. Their function in structuring discourse depends mainly on whether old or new information is in focus, and whether the current turn continues in the same context space or returns to a previous context space.

utt26	s:	okay **so** we'll get to Corning at eight a.m. and then go to where?
utt28	u:	Dansville
utt29	s:	okay
utt30	u:	**and** then Avon
utt31	s:	okay **so** Dansville at nine and then to Avon at ten okay
utt32	u:	**and** what time is it now?
utt33	s:	it'll be ten a.m. by the time that we get those **so** we should have two tankers of orange juice made right?
utt35	u:	okay

Ex. d93-10.3: Contrast of **and** vs. **so**

Notice that the 'user' continues his prior turn in utt30 with 'and', but when the system summarizes the running plan, he has to use 'so'. If he had said 'and Dansville at nine and then to Avon...' it would have implied that the upcoming utterance would build onto the plan.

3 ADJACENCY PAIRS AND DISCOURSE MARKERS

We have demonstrated above that a turn-initial discourse marker indicates the content of the upcoming speech. But the question remains, in what situations do speakers tend to begin an utterance with a discourse marker? Since discourse markers help speakers signal how the current turn relates to prior talk, we wanted to see what that prior talk tended to be. A traditional method for analyzing sequences of utterances is to organize them into adjacency pairs and to focus on their surface forms. Question/Answer pairs dominate the Trains dialogs, so the following traditional speech acts were used to annotate the dialogs:

Speech Acts

Acknowledge Backchannel 'Okay' or 'mm-hm'.

Check Restating old information to elicit a positive response from the partner (e.g. That was three hours to Bath?).

Confirm Restating old information, with no apparent intention of partner agreement.

Filled Pause A turn containing no information such as 'hm'.

Inform Information not previously made explicit.

Request Request for information.

Respond Respond to a Request.

Y/N Question Questions requiring a yes/no answer. Differ from Check because the speaker displays no bias toward which answer he expects.

Y/N Answer Answering 'yes', 'no', 'right', etc.

Table 2 shows the prior speech act of turns beginning with discourse markers. After an adjacency pair has been initiated, such as by a Y/N Question or Request Info, the next turn almost never begins with a discourse marker. The turn following the initiation of an adjacency pair is expected to be the second pair part. Since the role of that turn is not ambiguous, it will not begin with a discourse marker. It would indeed be odd if after a direct question such as 'So how many hours is it from Avon to Dansville' the system responded 'And 6' or 'So 6'. A possible exception would be to begin with 'well' if the upcoming utterance is a correction rather than an answer.

After a turn which is not the initiation of an adjacency pair, such as Acknowledge, Respond, or Inform, the next turn has a much higher probability of beginning with a discourse marker. If the prior speech act concluded an adjacency pair, the role of the next statement is ambiguous, so a discourse marker is used to provide an expectation about its relationship to prior discourse. Discourse markers, then, are used at the beginning of a turn which is structurally ambiguous.

	Total Turns	Turns beginning with				DM Turns
		And	Oh	So	Well	% of Total
Prior speech acts that initiate adjacency pair						
Check	23	0	0	0	1	4%
Request Info	45	0	0	1	0	2%
Y/N Question	8	0	0	0	0	0%
Prior speech acts that conclude adjacency pair						
Respond	38	3	2	5	1	30%
Y/N Answer	26	1	1	1	0	12%
Acknowledge	107	21	4	16	2	40%
Prior speech acts not in adjacency pair						
Confirm	42	2	0	0	1	7%
Inform	96	1	10	5	2	19%
Filled Pause	6	0	0	0	0	0%

Table 2: Prior speech act of DM-initial turns

4 GROUNDING AND DISCOURSE MARKERS

Spoken dialog involves a collaboration between two conversational agents. This collaboration is necessary to ensure that the agents maintain a coordinated set of beliefs about what each other is contributing to the dialog as well as the current state of the dialog. This is achieved by a process called *grounding* [1], and accounts for a significant amount of the conversational energy in human-human dialogs, as displayed by the frequent use of acknowledgments, back-channel responses, and paraphrases [5].

Since discourse markers signal how the current turn ties in structurally with the preceding dialog, do they also signal anything about the grounding of information provided in that turn? The same set of Trains dialogs were annotated using 6 strength levels of grounding evidence described below, and taken from [1]. The grounding level does not take into account any initial discourse markers.

Strength of Grounding Evidence

Level 0 The first turn in a dialog (which has nothing to ground) and filled pauses.

Level 1 'Continued attention' - Empty utterances

Level 2 'Next relevant contribution'- there is nothing in the turn which explicitly signals understanding of the other speaker's turn. (e.g. u: 'how many hours from Avon to Bath?' s: 'six hours' u: 'and from Bath to Dansville?'. The third turn would be marked level 2.)

Level 3 Acknowledgments

Level 4 'Demonstrated understanding' - explicit demonstration that the speaker understood the prior turn. (e.g. answers to questions, corrections)

Level 5 Verbatim repeat of part of the other speaker's turn

Level 6 Other-completions

The distribution of strength of grounding evidence for each discourse marker is listed in Table 3. The different discourse markers exhibit some interesting differences in terms of grounding strength. Turns that began with 'and' typically did not explicitly signal very strong grounding, in fact they typically did not explicitly register any acknowledgment of the other speaker's prior turn. This behavior makes sense in light of the discussion above of 'and's role to signal that the current turn is a continuation of the same speaker's prior turn. 'And' usually only marks grounding in the sense of appropriate next contribution.

Schiffrin discovered that using 'and' actually makes a claim that the speaker does not wish to cooperate with a potential topic shift introduced by the other speaker's intervening turn [8]. Turn-initial 'and' sends a strong signal that the current turn continues to build on information from the same speaker's prior speech. Look again at the example from dialog 10.3 in section 2.5. In utt31, the 'system' summarizes the plan and states the current time. But the 'user' is not paying attention and his next turn is "and what time is it now". The 'and' clearly marks this turn as a continuation of the train of thought started in his previous turn.

utt12	s:	oh I should tell you that you can only pull three loaded boxcars at a time
utt13	u:	oh okay **so** why don't we go from Avon to Bath, get the two boxcars
utt15	s:	okay
utt16	u:	**and** go back up and get bananas and...

Ex. d93-17.2: Grounding behavior of **and** vs. **so**

'So', on the other hand, tends to explicitly mark that the speaker is taking into consideration information that was just provided by the other speaker. A summary building on information provided by the other speaker, as in utt13 above, is typically marked with 'so'.

Since the discourse function of 'oh' and 'well' revolve around their speaker's orientation to information, they are expected to carry higher grounding strength. Indeed, both are above the dialog average and 'well' is the highest of all.

The average for non-discourse-marker turns is 3.06, but for DM-initial turns the average is 2.49. The average for non-discourse-marker turns is pulled up by the high occurrence of responses, which carry level 4 grounding. As we saw in the prior section, responses to questions rarely begin with discourse markers. Discourse markers, then, are used to help signal grounding in turns that would lack clear grounding evidence without the discourse marker.

Grounding Strength	All Turns	Turns beginning with				DM Turns % of Total
		And	Oh	So	Well	
Level 0	14	1	0	0	0	7%
Level 1	24	0	0	0	0	0%
Level 2	113	26	9	19	3	50%
Level 3	139	0	5	1	0	4%
Level 4	72	1	2	6	4	18%
Level 5	21	0	1	2	0	14%
Level 6	6	0	0	0	0	0%
Average	2.56	2	2.71	2.68	3.14	

Table 3: Grounding strength in DM & non-DM turns

Turns which demonstrate grounding more explicitly, such as responses, tend not to carry an initial discourse marker.

5 CONCLUSION

Our analysis shows that discourse markers provide an important first clue to the conversational move about to be made by a speaker. They also provide important information about the current speaker's orientation to information provided in the prior turn, so they should be exploited by dialog systems to coordinate shared beliefs. Current spoken dialog systems tend not to utilize this important source of information in understanding the user's utterance. Our results also have implications for utterance generation in task-oriented systems. To collaborate in the same way humans do, our systems must construct utterances which signal mutual understanding of shared information and discourse structure in an appropriate way. Using discourse markers will help the system's utterances seem more natural to the user, and will help him understand the discourse intent of the upcoming move.

6 REFERENCES

[1] Herbert H. Clark and Edward F. Schaefer. Contributing to discourse. *Cognitive Science*, 13:259–294, 1989.

[2] Robin Cohen. A computational theory of the function of clue words in argument understanding. In *Proceedings of the 10^{th} International Conference on Computational Linguistics (COLING '84)*, pages 251–255, 1984.

[3] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[4] Peter A. Heeman and James F. Allen. The Trains spoken dialog corpus. CD-ROM, Linguistics Data Consortium, April 1995.

[5] Peter A. Heeman and David Traum. Utterance Units and Grounding in Spoken Dialogue. In *Proceedings of the 4^{th} International Conference on Spoken Language Processing*, pages 1884-1887, 1996.

[6] B. J. Heritage. A change-of-state token and aspects of its sequential placement. In *Structures of social action: Studies in conversation analysis*, pages 299-345. Cambridge University Press, 1984.

[7] Rachel Reichman. *Getting Computers to Talk Like You and Me*. MIT Press, Cambridge, Mass., 1985.

[8] Deborah Schiffrin. *Discourse Markers*. Cambridge University Press, New York, 1987.