Towards Understanding Mixed-Initiative in Task-Oriented Dialogues

Fan Yang, Peter A. Heeman, Kristy Hollingshead

Center for Spoken Language Understanding
OGI School of Science & Engineering
Oregon Health & Science University
`{fly,heeman,hollingk}@cslu.ogi.edu`

Abstract

Mixed-initiative in dialogue does not mean that either speaker is free to take the lead in the conversation at any time. Rather, who can show initiative, and when, is restricted. In this paper, we define initiative and control as two levels of dialogue phenomena, and give further evidence for our theory of restricted initiative that control belongs to the initiator of a discourse segment. We show that the initiator and the non-initiator play different roles in terms of showing initiative.

1. Introduction

Initiative involves actively contributing to a dialogue. Researchers are interested in building mixed-initiative dialogue systems that users can interact with as easily as they interact with other people. On one hand, a dialogue system should allow human users to actively participate in the reasoning procedure; for example, to ask a question or to make a proposal. On the other hand, a dialogue system also needs to actively contribute to a dialogue, efficiently and effectively helping human users to accomplish a task. Thus, a dialogue system should know when to take the lead and when to let the human user take the lead. There are conventions in human-human dialogues that govern this behavior. Such conventions help guarantee an interaction mode that is convenient to human users and supports human-style problem solving.

A number of theories have been proposed for how initiative is managed. Whittaker and Stenton [1] proposed that initiative does not simply bounce back and forth, but that once one person has it, the person tends to keep it for a while. That is, the person who shows initiative is very likely to show the next initiative. Chu-Carroll and Brown [2] proposed that there are two levels of initiative, with dialogue initiative subservient to task initiative. Guinn [3] and Allen [4] proposed that whoever is most capable of the next (sub-)task takes initiative. Cohen et al. [5] proposed four different models of initiative, the most elaborate being where a speaker has initiative in a process; and they also proposed different degrees of initiative. However, none of these theories analyzed initiative from the viewpoint of discourse structure, or tied it to the conversational goals of the participants.

In our previous work [6, 7], we analyzed a subset of the Trains dialogues [8] using Grosz and Sidner's discourse structure theory [9] and annotated them as to which utterances demonstrate initiative. We found that initiative tends to be mainly shown by the speaker who initiated a discourse segment. When the other speaker, the non-initiator, shows initiative, either by making an utterance or a subsegment, the initiator tends to be the one who shows initiative next. We used this as evidence that initiative is subordinate to discourse structure.

This paper expands on our previous work in the following ways. First, we clarify the definitions of initiative and control, and we restate our theory in terms of these definitions. Second, we give further evidence that control is subservient to discourse structure. We re-evaluate our past results to take into account the different roles of the two participants in the Trains corpus, the system and the user. We also look at the patterns of initiative.

2. Redefining Initiative and Control

Most of the work on defining mixed-initiative has used the terms control and initiative interchangeably. These terms seem to be used at two separate levels. The first is at the utterance level. When a speaker makes a certain type of utterance, the speaker is said to be showing control or initiative. The second is at a more global level, and is something that the speaker has that lives on beyond individual utterances. We treat these two phenomena as separate, and use initiative to refer to the first, and control to the second.[1]

Definition 1: *Initiative* is the behavior of volunteering new information without being explicitly requested to, or of asking a question to establish mutual belief, or of proposing a new domain goal to achieve.

---

[1]Cohen et al. [5] used these two terms opposite from our use of them.

Initiative is short-lived. Initiative begins and ends with the utterance. In the dialogue excerpt in Figure 1, the user shows initiative in utterance u1, but nobody shows initiative in utterance u2.

| user | u1 | I need to get a boxcar at Avon |
| system | u2 | ok |

Figure 1: An example for initiative and control

Definition 2: *Control* is the authority to drive a dialogue (or sub-dialogue) towards the achievement of its goal.

In contrast to initiative, control has lasting effect. Once a speaker has control, he or she will have it until the other speaker has it.

Our definition of control does not specify its relation to initiative. This is what a theory of mixed-initiative must do. The theory must also specify how and when control changes. According to Whittaker and Stenton's theory [1], when a speaker shows initiative, this gives control to that speaker, and the speaker keeps control until the other speaker shows initiative. For the examples in Figure 1, since the user shows initiative in u1 and the system does not show initiative in u2, the user is said to have control in u1 and u2.

According to our theory of mixed-initiative [6, 7], control follows discourse structure. The speaker, say A, who initiated the discourse segment is said to be in control. If the non-initiator, B, makes an utterance that shows initiative, this does not give B control. The same is true when B makes an utterance that starts an embedded discourse segment. B has control for the extent of the embedded discourse segment, but once that segment finishes, control reverts to A. Grosz and Sidner [9] tied discourse segments with conversational goals; hence, when a speaker proposes a conversational goal, that speaker is in control over the extent of the conversation to achieve that goal.

In order for our theory of mixed-initiative to be sound, there must be differences in behavior between the initiator of a discourse segment and the non-initiator. In other words, the 'authority to drive a dialogue' must have some manifestation. In our previous work [6, 7], we showed that the initiator tends to show initiative 8.3 times more often than the non-initiator and that after the non-initiator's utterance of initiative or subsegments, the next person to show initiative tends to be the initiator of the current discourse segment.

3. Dialogue Annotations

We used the same annotations as were used in [7]. We annotated hierarchical discourse segments according to Grosz and Sidner [9]. In deciding whether a group of utterances formed a discourse segment, we took into account whether there existed a shared and recognizable goal. We also followed Carletta et al. [10] by distinguishing two types of segments: transaction segments and dialogue games. Transaction segments typically correspond to sub-tasks, such as stating the goal of the plan, constructing the steps in the plan, refining the plan, summarizing the plan and evaluating the plan. Dialogue games are related to adjacency pairs, which typically consists of a first part and a second part. Each part can have one or more utterances. We explicitly mark the second part as a special type of segment. Second parts are special in that control should be viewed as explicitly passed to the non-initiator of the dialogue game. In a dialogue game, the initiator gives up control to the non-initiator after the first part. The non-initiator of the dialogue game takes over the control when initiating the second part, and relinquishes the control when completing the segment. We call such phenomena *delegating control*. We did not analyze what happens inside a second part. Here we just assume that the second part as a whole does not show initiative. However, sometimes the non-initiator of a dialogue game might over-answer questions; sometimes the initiator of the dialogue game might even show initiative inside a second part. These are interesting behavior that need further research.

We did not directly code which utterances show initiative. Instead, we tagged each utterance with a variation of the DAMSL scheme [11]. Utterances were tagged as forward or backward functions, or others. Forward functions include statements, questions, checks and suggestions. Backward functions include agreements, answers, acknowledgments, repetitions and completions. Others include stalls and move-ons. We used utterance tags to derive whether an utterance shows initiative. Utterances coded as forward functions show initiative while backward functions and others do not.

We annotated twelve dialogues totaling 50 minutes from the Train corpus using DialogueView [12]. Two expert annotators annotated each dialogue for utterance tags and discourse structure. They then compared their annotations to reach a consensus using ACT [13]. Most of the differences between the annotators were resolved successfully.

4. Evidence

In this section, we build on our previous work to give further evidence for our theory of mixed-initiative interaction.

4.1. Frequency of Initiative

In our previous work [6, 7], we reported on how frequently the initiator and non-initiator of a discourse segment show initiative. However, we did not distinguish between the two roles that the speakers play.

In the Trains dialogues, one speaker is playing the role of a user and the other is playing the role of the sys-

Table 1: Frequency of Initiative

	Initiator:	
	System	User
Initiator shows initiative		
in first utterance	66	145
in other utterances	108	302
Non-initiator shows initiative	8	43

Table 2: After a speaker shows initiative

	Initiator:	
	System	User
Initiator shows initiative followed by		
End of Segment	30	66
Initiator shows initiative	120	381
Non-initiator shows initiative	13	46
Ratio of Initiator to non-initiator	9.2	8.3
Non-Initiator shows initiative followed by		
End of Segment	8	16
Initiator shows initiative	6	50
Non-initiator shows initiative	3	22
Ratio of Initiator to non-initiator	2.0	2.3

tem. The user has to come up with a plan to manufacture and ship goods to achieve a certain goal. The system has complete knowledge of the domain, such as the distances between cities, how long it takes to manufacture the goods, and constrains on how much cargo a train can pull. The system also has access to pen and paper. The system is told not to be 'too smart' and not to solve the problem for the user, but also not to be too limited [8].

Of the 233 transaction segments and dialogue games in our annotated dialogues, 69 were initiated by the system and 164 by the user. Table 1 reports the number of utterances made by the initiator and non-initiator that show initiative. Column 2 gives the results where the system is the initiator, and column 3 gives the results where the user is the initiator.

We separate the first utterance in a segment, as by definition, it is always made by the initiator. In discourse segments initiated by the system, the initiator (system) shows initiative 13.5 (108/8) times more often than the non-initiator (user). In discourse segments initiated by the user, the initiator (user) shows initiative 7 (302/43) times more often than the non-initiator (system). Even though the roles of the system and user are so different, we see that both behave similarly when acting as the initiator of a discourse segment. This supports our claim that the initiator is the one who is driving the conversation, and hence is in control.

4.2. After Non-initiator Shows Initiative

Our model of mixed-initiative allows for the non-initiator to make an utterance that shows initiative, and to even initiate a subsegment. In our previous work [7], we analyzed what happens after the utterance or the subsegment. We found that the next speaker to show initiative tends to be the initiator. This result contradicts Whittaker and Stenton's theory of mixed-initiative, as they would predict that when the non-initiator shows initiative, the non-initiator would take over control and hence be very likely to show initiative next.

In this subsection, we repeat our previous analysis, but distinguish between the roles that the speakers play in the dialogues. Table 2 gives the results. Column 2 is where the system is the initiator and column 3 is where the user is.

First, consider the case where the system is the initiator of a discourse segment. After the initiator shows initiative (by making an utterance, or initiating a subsegment), the next person to show initiative (after that utterance or the subsegment) is the initiator in a ratio of 9.2:1 (120/13). In those cases where the non-initiator shows initiative, a ratio of 2:1 (6/3), the initiator is the next to show initiative. Note that these two ratios are almost identical to the ratios when the user is the initiator (8.3:1 and 2.3:1 respectively). Hence, even though the roles played by the system and the user in the dialogues are very different, the way they behave when they are the initiator is very similar.

4.3. Initiative Patterns

The previous subsection showed that when the non-initiator shows initiative, the next person to show initiative is bound to be the initiator. However, in 25 of 81 cases, the non-initiator shows initiative at least twice in a row. To understand these cases, we looked at patterns of initiative in discourse segments.

To determine the pattern of initiative in a discourse segment, we carry out the following steps. First, any embedded segments are viewed as an utterance that shows initiative by the speaker who initiates the subsegment. Second, we remove any utterances that do not show initiative. Third, we count up how many initiative utterances in a row each speaker alternately makes. Thus the pattern "3-2-4-1" represents a discourse segment in which the initiator makes 3 utterances (or subsegments) that show initiative, followed by 2 by the non-initiator, followed by 4 by the initiator and ending with 1 by the non-initiator. Hence, the initiator shows initiative 7 times in the segment while the non-initiator shows initiative three times.

We classify initiative patterns into three types: good patterns, bad patterns, and questionable patterns. Good patterns are patterns that satisfy the following conditions:
Majority: initiator shows at least 80% of the initiative.
Example: 5-1 Counterexample: 3-1
Isolated: non-initiator never shows initiative more than once in a row.
Example: 4-1-3-1 Counterexample: 4-2-3-1

Non-alternating: no subpattern of 1-1-1 (starting with non-initiator).

Example: 8-1-3-1   Counterexample: 8-1-1-1

We feel that good patterns are clear examples where the initiator is driving the conversation. A pattern is *bad* if the initiator shows control less than 60% of the time. For bad patterns, it is not clear that the initiator is indeed driving the conversation. All other patterns are viewed as questionable. In a segment with a questionable pattern, the initiator tends to be controlling (showing most cases of initiative), yet there might be local extent where it is unclear who is the controller.

Table 3: Patterns of Initiative

	Initiator:	
	System	User
Good		
Only initiator shows initiative	17	25
Others	5	15
Questionable	0	11
Bad	2	8

In our annotated dialogues, we focused on the 83 transaction segments. The results are reported in Table 3. In total we see that 75% of the segments have a good initiative pattern. More research work is needed to understand what is happening in segments with the other two patterns. It is also interesting to note that the majority of these segments were initiated by the user. These segments might be ones in which heavy collaboration is occurring in order to solve a difficult subproblem, and might correspond to the shared turns of Shriffin [14]. Further work is needed to investigate these segments.

5. Conclusion

In this paper, we give a clear definition of initiative and control. We show that the initiator and the non-initiator play different roles in terms of showing initiative. The initiator tends to show initiative continuously, with the non-initiator interrupting occasionally. Control belongs to the initiator. This proposal has important implications for dialogue management, as it will pave the way for building dialogue systems that can engage in mixed-initiative dialogues. Mixed-initiative dialogues do not mean that both speakers show initiative freely without any restrictions. Instead, mixed-initiative means that speakers contribute at the most appropriate timing when it is best suited. Our theory points out that the initiator is the person who mainly shows initiative, while the non-initiator mainly just responds. Human language has evolved over several generations and reached some local optimization. Perhaps people adopt such convention just because conversation in this way is more efficient. Our future work also includes understanding the benefits of using this model in human-human dialogues and in building dialogue systems.

6. References

[1] S. Whittaker and P. Stenton, "Cues and control in expert-client dialogue," in *Proceedings of 28th ACL*, 1988.

[2] J. Chu-Carroll and M. K. Brown, "An evidential model for tracking initiative in collaborative dialogue interactions," *User Modeling and User Adapted Interaction*, vol. 8, pp. 215–253, 1998.

[3] C. I. Guinn, "An analysis of initiative selection in collaborative task-oriented discourse," *User Modeling and User Adapted Interaction*, vol. 8, pp. 255–314, 1998.

[4] J. F. Allen, "Mixed-initiative interaction," *IEEE Intelligent Systems*, September-October 1999.

[5] R. Cohen, C. Allaby, C. Cumbaa, M. Fitzgerald, K. Ho, B. Hui, C. Latulipe, F. Lu, N. Moussa, D. Pooley, A. Qian, and S. Siddiqi, "What is initiative?" *User Modeling and User Adapted Interaction*, vol. 8, pp. 171–214, 1998.

[6] S. E. Strayer, P. A. Heeman, and F. Yang, "Reconciling control and discourse structure," in *Current and New Directions in Discourse and Dialogue*, J. van Kuppevelt and R. Smith, Eds. Kluwer, 2003.

[7] P. A. Heeman, F. Yang, and S. E. Strayer, "Control in task-oriented dialogues," in *Proceedings of 8th eurospeech*, September 2003.

[8] P. A. Heeman and J. Allen, "The Trains 93 dialogues," Department of Computer Science, University of Rochester, Trains Technical Note 94-2, March 1995.

[9] B. J. Grosz and C. L. Sidner, "Attention, intentions, and the structure of discourse," *Computational Linguistics*, vol. 12, no. 3, 1986.

[10] J. Carletta, A. Isard, S. Isard, J. C. Kowtko, G. Doherty-Sneddon, and A. H. Anderson, "The reliability of a dialogue structure coding scheme," *Computational Linguistics*, vol. 23, no. 1, 1997.

[11] M. G. Core and J. F. Allen, "Coding dialogues with the DAMSL annotation scheme," in *Proceedings of AAAI Fall 1997 Symposium*, 1997.

[12] P. A. Heeman, F. Yang, and S. E. Strayer, "DialogueView: A dialogue annotation tool," in *Proceedings of 3rd SIGdial workshop on Dialogue and Discourse*, Philadelphia, 2002.

[13] F. Yang, S. E. Strayer, and P. A. Heeman, "ACT: A graphical dialogue annotation comparison tool," in *Proceedings of 7th ICSLP*, Denver, 2002.

[14] D. Schiffrin, *Discourse Markers*. Cambridge University Press, 1987.