How do people construct mutual beliefs in task-oriented dialogues?

Yoshiko Kawabata^{*†} and Toshihiko Matsuka^{*} * Chiba University, Chiba, Japan E-mail: ykawabat@gmail.com Tel/Fax: +81-043-290-2277 † National Institute for Japanese Language and Linguistics, Japan

Abstract—The present study investigates how mutual beliefs are achieved by examining the relationship between actual behaviors and utterances in task-oriented dialogues. According to a widely accepted model, mutual belief about a task is considered to be achieved when a listener accepted utterances about the task given by another agent and gives some signs of task completion to the agent. However, by analyzing Japanese Map Task Dialogue Corpus (JMTDC), we found vast majority of conversations (94%) did not follow what was suggested by the model. We categorized those non-standard dialogues into six categories, namely, *delayed acceptance, premature sign of completion, execution postponement, silent adjustment, unconfirmed*, and *indirection*. We further analyzed those six categories carefully to see how and when participants were able to achieve mutual belief in the dialogues.

I. INTRODUCTION

A successful dialogue usually indicates a dialogue where the listeners are able to infer intentions and functions of individual utterances of speakers, such as speakers' meanings [1] and illocutionary acts [2][3]. However, what matters the most in dialogue is whether the objective of a dialogue is achieved or not. We often engage in dialogues to achieve various objectives and purposes such as lectures at universities, making reservations, arranging a meeting with friends. In these activities, correctly understanding the speaker's intention in individual utterances may be necessary in the sense that it contributes to the achievement of its ultimate goal. One crucial factor that contributes to achieving the goal is that everyone in the conversation constructs the mutual belief about the goal; what a professor is intended to convey is correctly understood by students; a travel agents correctly understand his or her customer's intended itinerary; or you and your friends correctly understand the time and place to meet.

Understanding how people construct a mutual belief through conversation is very important not only because it leads to better understandings of the nature of humans' conversational activities but also it helps us develop an intelligent interactive robot and/or interface that cooperate and collaborate with humans to perform various tasks. The present paper tries to examine the mechanisms and processes that people employ to construct mutual beliefs.

However, it is difficult to infer what listeners understand by simply examining what is said in conversations, which in turn make us difficult to investigate the mechanisms with which people construct mutual beliefs using ordinary conversations. To overcome this difficulty, we analyzed activities and conversations in Japanese Map Task Dialogue Corpus (JMTDC) [4] in which pairs of agents collaboratively performed tasks in separate location without any visual feedback of others' actions. JMTDC was created with reference to HCRC Map Task Dialogue Corpus [5] using Japanese speaking participants. There were two participants in each corpus data set. In each data set, one participant verbally gave directions (we refer to this type of participants as a "giver" hereafter) to the other participant (i.e., follower) who was instructed to draw the given route. Although the maps that were given to the givers and followers were roughly the same, they were slightly different (see Fig. 1).

JMTDC is particularly suitable to our study in the following two points. First, what the followers understand and interpret would become explicitly apparent as they were instructed to draw what were said by givers. Second, slightly different maps and absence of visual feedback on actions would make tasks and situations more uncertain in which in turn would generate more misunderstandings and misinterpretations, promoting actions to seek construction of mutual beliefs. It should be noted that all participants were informed that there might have been slight differences between maps provided to the givers and followers. By analyzing dialogues and actions in JMTDC, the present study examined how people construct mutual belief through conversation. In particular, we paid close attention to the followers' actions (i.e., drawing routes) and their relationships with conversations.

II. THEORETICAL BACKGROUND AND HYPOTHESIS

According to Theory of Discourse Structure by Grosz and Sidner [6], the main discourse purpose (DP) of participating a conversation consists of multiple segmented purposes (discourse segment purpose or DSP) that lead to achievement of DP. DSP is comparable to discourse segment (DS) which consists of a collection of utterances. Considering map task dialogues according to Theory of Discourse Structure, DP is equivalent to drawing a whole route from the starting point to the final destination. In map task dialogues, givers generally do not give descriptions of a whole route to their followers at once. Givers usually divide the whole route into several segmented routes, and give descriptions of the segmented routes to followers iteratively [7]. Thus, description of a segmented route can be considered as DSP and sequence of DSP as a segmentation of DS.



Fig. 1. Maps given to participants in Japanese Map Task Dialogue Corpus. (a) Giver's map. The route was drawn by the experimenter before the task began. (b) Follower's map. The route was drawn by the follower during the experiment.

A. Traum's model

One of the most successful models for task-oriented conversation is that of Traum's [8], and we used it as the reference model in the present paper. Traum's model was developed by integrating models proposed in the fields of Artificial Intelligence (e.g., speech act) [9][10][11] and Psycholinguistics (e.g. conversational grounding) [12].

In map tasks, given that only a giver knows the correct route, a conversation about a description of (or a request to draw) a segmented route, say point P_1 to point P_2 is predicted to be initiated by the giver. When a follower understand a request given by a giver, then the follower is predicted to give some sign of acceptance of the request. By saying and hearing the sign of acceptance, both follower and giver share a belief that drawing the route P_1 to P_2 is planned and will be executed by the follower in near future, constructing mutual belief. We refer this sort of mutual belief about an event and/or situation that is going to happen to as prospective mutual belief (P-MB). When the follower draw a segmented route (i.e., executing the request), then DSP becomes satisfied. However, in map tasks, a giver never has a chance to directly observe what his or her follower had drawn or done. Thus, the follower is predicted to confirm that the task or DSP being completed by giving some sign of completion to the giver [10]. We refer this sort of mutual belief about an event and/or situation that is already happened to as retrospective mutual belief (R-MB). This sequence of utterances and actions is depicted in Fig. 2. One DSP in a map task starts with (a) a request (to draw a segmented route), followed by (b) acceptance (of the request), (c) execution (of the request), (d) sign of completion, and (e) a next request indicating a transition to next DSP.

III. DATA

We analyzed Japanese Map Task Dialogue Corpus (JMTDC) in order to examine how mutual beliefs were constructed. There were 64 students (32 males and 32 females)



Fig. 2. Predicted sequence of conversational and behavioral actions by Traum's Model

from Chiba University participated in JMTDC experiment. They participated in the experiment in a pair with an acquaintance. Each pair were also paired with another pair to form a group of four participants. Each participant completed the map tasks four times (twice as a giver and twice as a follower with different individuals). A total of 128 dialogues were recorded. The half of the 128 dialogues were recorded under a situation in which the subjects were able to see each other's face and the other half was without eye contact. In the present study, we only analyzed the latter half as we wanted to create more uncertain situations so that participants were more likely to exhibit actions to construct mutual beliefs. Among the 64 non-eye contact dialogues, we analyzed randomly selected four dialogues.

We, first, annotated the followers' action using ELAN [13]. Followers' action were categorized into four categories, namely *route drawing*, *route tracing* (i.e., tracing a route that was already drawn), writing down a land mark, and marking (e.g., marking a targeted point or incorrect route). Among the four categories, we extracted dialogues and actions associated with *route drawing* as it was most likely to be associated with the discourse purpose (DP). In particular, for each *route drawing* of all followers, we extracted five core actions, namely (a) request, (b) acceptance, (c) execution, (d) sign of completion, and (e) next request, if there were any. As we will describe later, some of the above except execution were absent in some dialogues. We then categorized each dialogue on the basis of how the five actions were ordered.

IV. RESULT

Among the four dialogues, we found several types of discourse segment (DS), both Standard DSs that were well expected by Traum's model and non-standard DSs that did not follow the model.

A. Standard Discourse Segment

Dialogues where the five core actions were arranged as in Fig. 2 were categorized as standard Discourse Segment (DS). An example of standard DS is show in Fig. 3. Although, we expected a majority of dialogues would be categorized as standard DS, only six percent of dialogues were this type.

B. Non-Standard Discourse Segment

The remaining 94 percent of dialogue were categorized as non-standard DS. We further categorized non-standard DS into



Fig. 3. Excerpt of standard discourse segment (DS). Numbers in each column indicate the times of beginning and end of corresponding utterance. Letters G (or F) next to the numbers indicate speaker of corresponding utterance was giver (or follower). All conversations were done in Japanese language. Their English translation is written in litalics

six categories. The causes of irregularities in four of nonstandard DS could be attributed to followers, and the other two were caused with givers.

1) Delayed Acceptance: According to Traum's model, in collaborative tasks, members are assumed to agree with each other about their tasks and share their plans to achieve the (sub) goal of the task. However, there are some cases, follower execute a request (drawing a route) before giving some sign of acceptance (Fig. 4). We call this non-standard DS as delayed acceptance. An example of delayed acceptance is shown in Fig. 5. One important function of acceptance (of request) is to achieve prospective mutual belief (P-MB: mutual belief about an event and/or situation that is going to happen). However, in delayed acceptance, while acceptance would be interpreted as an utterance to construct P-MB by givers, but it would not be meant to construct P-MB to followers.



Fig. 4. Sequence of actions in delayed acceptance

2) Premature Sign of Completion: While a sign of acceptance helps agents in dialogue to construct prospective mutual belief, a sign of completion helps them construct retrospective mutual belief (R-MB: mutual belief about an event and/or situation that is already happened). One type of inappropriate usage of sign of completion is *premature sign of completion* where a sign of completion is given before completing tasks (Fig. 6). Excerpt of premature sign of completion is shown in Fig. 7. The follower in this dialogue said "I wen to the right." before actually arriving to the location he or she was requested



- I arrived there. [sign of completion]

Fig. 5. Excerpt of delayed acceptance

to be. Thus, like delayed acceptance, while sign of completion would be interpreted as an utterance to construct R-MB, but it was not be meant to construct to R-MB to followers.



Fig. 6. Sequence of actions in premature sign of completion

(1) 00:29:677-00:36:449 G: daitai sita-kara daitai yaku 1 senti kurai-no tokoro-no tokoro-de use eto migi-ni magatt-te-kudasai

At approximately from 1cm from the bottom, turn to the right.[request]

- (2) 00:33:952-00:36:816 F: hai a hai Yes. uh yes. [acceptance]
- (3) 00:37:032-00:37:379 G: migi-ni hai To the right. Yes.
- (4) F: (started writing route) [initiated execution]
- (5) 00:37:526-00:38:753 F: hai hai migi iki-masi-ta Yes. Yes. I went to the right. [sign of completion]
- (6) F: (wrote route) [finished execution]

Fig. 7. Excerpt of premature sign of completion

3) Execution postponement: There are some dialogues in which followers gave signs of completion even when they had not initiated any action for what they were requested and accepted to do so (Fig. 8). We call this type of DS as execution postponement. Execution postponement is another type of inappropriate usage of sign of completion, like premature sign of completion. In both premature sign of acceptance

and execution postponement, sign of completion was not be meant to construct to R-MB to followers. However, while followers had initiated actions and completions of the actions were reasonably expected (by followers) in premature sign of completion, followers had not initiated any action and there were greater levels of uncertainly whether requested tasks would be completed or not in execution postponement.



Fig. 8. Sequence of actions in execution postponement

| (1) | 03:35:638-03:44:165 G: sono Haioku-no hidariue-n- |
|-----|--|
| | toko-kara daitai yaku 5 senti gurai eto sono 45 do gurai |
| | sagatt-te-kudasai |
| | Go down about 5cm at 45 degrees from the Deserted- |
| | house. [next request] |
| (2) | 03:41:168-03:41:376 F: hai |
| | Yes. [acceptance of next request] |
| (3) | 03:42:790-03:44:050 F: (wrote route) |

[execution of *previous* request]

Fig. 9. Excerpt of execution postponement

4) Silent adjustment: There were some cases where followers noticed that some segments of the route that he or she had already drawn were incorrect and then corrected them without letting givers know about mis-drawings and corrections. We call this phenomenon silent adjustment. DS that resulted in mis-drawing might have done in a form of standard DS, and thus silent adjustment may not be considered as a non-standard DS. But, in terms of mutual belief, silently adjusting mistakes causes a delay in updating mutual belief by givers and followers. In silent adjustment, while a giver would think that he or she has achieved R-MB when s/he hears a sign of completion, a follower re-achieves R-MB when he or she corrects a mistake. If a follower tells about a mistake and its correction to a giver, then they re-achieve R-MB simultaneously. In this sense, we consider silent adjustment as a type of non-standard DS.

There is one important difference between silent adjustment and previous three non-standard DSs caused of followers. While previous three non-standard DS were deliberately made by followers (i.e., the followers in those non-standard DS had a chance to talk to the giver in a form of standard DS, but did not), the followers in silent adjustment resulted in a nonstandard DS not deliberately. In other words, the followers in previous three non-standard DSs deliberately caused some delays in achieving mutual belief or might have considered delays in mutual beliefs were not crucial. 5) Unconfirmed: There were some cases where DS resulted in a form of non-standard that were caused or initiated givers. For example, there were some cases in which givers started talking about next request without hearing a sign of completion from followers (Fig. 10). The givers in JMTDC never had a chance to see what the followers had drawn. Thus they could not have confirmatory evidence whether DSPs had achieved or not without hearing signs of completion by followers. We call this type of non-standard DS as *unconfirmed*.

Excerpt of unconfirmed is shown in Fig. 11. In this excerpt, the follower's acceptance (i.e, 2) allowed the giver and follower to construct P-MB, but the giver initiated next request (i.e., 5) without hearing any sign of completion or constructing R-MB.



Fig. 10. Sequence of actions in unconfirmed

| (1) | 06:26:346-06:29:196 G: sono kawa eto numa-ni sott-te |
|-----|--|
| | massugu itt-te-kudasai |
| | Go straight along the ponds. [request] |
| (2) | 06:27:984-06:30:518 F: hai hi hidari-ni hai |
| | Yes. To the left. Yes. [acceptance] |
| (3) | 06:30:112-06:30:256 G: hai |
| | Yes. |
| (4) | 06:30:425-06:32:985 F: (wrote route) |
| | [execution] |
| (5) | 06:31:520-06:32:978 G: so-si-tara mokuhyoo-titen-tte |
| | ari-mase-n-ka |
| | Then go to the goal around there [next request] |
| | Then, so to the sour around there. [next request] |

Fig. 11. Example of unconfirmed

6) Indirection: Another type of of non-standard DS caused or initiated by givers is *indirection*. There were several cases where givers' utterances were not in a form of request nor instruction, but contained information about route. The followers took or tried to take the utterances that contained information about routes but were not requested or instructed to draw as a form of request in these cases. We refer DSs that contained this sort of "request" to as *indirection*, because utterances were interpreted as requests in an indirect manner.

Fig. 12 shows an example of indirection. The giver originally stated a hypothetical moves or route in (1) and (3), and kept stating further hypothetical move (i.e, 5). At this moment, the follower noticed that the movement or route stated in (1) was meant to be drawn by him or her. The lack of (formal) request caused the follower not to provide any sign of acceptance of request simply because there was no request. The sequence of core action in indirection is shown in Fig. 13. Given that acceptance is missing in indirection, givers may consider that constructing P-MB is not important and follower usually do not have a chance to construct P-MB when DS is in a form of indirection.

- (1) 07:42:080-07:47:736 G: Marui-iwa-kara sonomama migi-ni dasu-to mizuumi-no hottan-to iu-ka kou hidari-ni tuki-masi-ta-ne *If you went right from the Round-rock, you have reached the tip of the pond, right?* (2) 07:46:704-07:48:496 F: a tyotyoi-tte deteru tokoro *Ah, the point of edge of the pond.*
- (3) 07:48:432-07:49:312 G: ni tuki-masu-yo-ne *You would reach there.*
- (4) 07:49:264-07:49:488 F: hai Yes
- (5) 07:50:064-07:51:898 G: tui-tara sore-o gurutto eto i If you reached there, go around it. [indirect request]
 (6) 07:51:790-07:53:181 F: a toriaezu zya tui-te
- *Oh, anyway I got there.* (7) 07:52:220-07:55:460 F:(wrote route)
 - [execution]





Fig. 13. Sequence of actions in *indirection*

V. DISCUSSION AND CONCLUSIONS

In *delayed acceptance*, followers started drawing route before constructing mutual belief with givers that they would draw the route in near future. In *premature sign of completion* and *execution postponement*, followers gave some sign of completion to givers even when they had not completed what were requested to do so. Thus, it seemed that followers found constructing prospective and retrospective mutual belief simultaneously with givers unimportant. In *unconfirmed*, givers gave another request even though follower had not given signs of completion. These conversational acts seem dishonest acts by speakers. Likewise, "requesting" followers without requesting to do so by givers (i.e., *indirection*) and postponing route drawing (*execution postponement*) by followers seem not rational nor productive. In this section we discuss reasons why those behaviors were exhibited by participants.

One reason is the nature of environments in JMTDC. In JMTDC, participants were not able to see others action, and thus givers never had a chance to know whether followers

drew routes before accepting requests nor whether followers gave signs of completion before finish drawing the routes. Thus, it was easier followers to exhibit delayed acceptance, premature sign of completion, and execution postponement in JMTDC environment. Unconfirmed might have also caused by the nature of JMTDC environment. In JMTDC, participants were explicitly instructed to draw the complete route by the experimenter, but they were not to drawn each segmented route iteratively. Thus, givers might have thought that acceptance of request was sufficient to infer that drawing segmented route would be reasonably completed by follower, and then moved to next DSP. Although JMTDC environment created many opportunities to participants to exhibit non-standard DS, it does not explain why these non-standard DSs actually occurred so frequently. We now discuss reason why nonstandard DSs actually occurred by particularly focusing on execution postponement and silent adjustment.

Figs. 14 and 15 shows another example of *silent adjustment*. The giver checked whether the current location of the follower was right (P_G in Fig. 15) by stating (1) in Fig 14. To the question, the follower gave an affirmative response by stating (2). But, after giving affirmative response, the follower silently adjust his or her location, from P_{F1} to P_{F2} . It is obvious that the follower's adjustment was triggered by the giver's utterance just before the adjustment. The follower was most likely to infer that the giver were giving another request by assumed that the follower was at P_{F2} , and he or she made the correction. Lewis [14] called the mechanism of inferring the premise of utterance as "The rule of accommodation for presupposition (accommodation rule, hereafter) In, particular Lew stated:

If at time t something is said that requires presupposition P to be acceptable, and if P is not presupposed just before t, then - ceteris paribus and within certain limits - presupposition P comes into existence at t. (p.340)

To understand the utterance of the giver, the follower needed to accept that he or she needed to be at P_{F2} and made an appropriate adjustment. If followers uses the accommodation rule, either explicitly or implicitly, they may hold drawing a route until the description of next segmented route becomes available like in execution postponement (Figs. 8 & 9). In JMTDC, the starting location of every segmented route is the destination of its preceding segmented route. Thus, the givers gave requests for route drawing by presuming that the followers were at the destinations of previous segmented route. By hearing the utterances about next request, the followers were able to infer the correct destination of the previous request when they applied the accommodation rule. This might have been the main reason why the followers deliberately postponed drawing routes they were requested to draw. Likewise, the followers silently adjusted their locations in order to accomplish the main goal of the task without letting the givers know that they were at incorrect locations, mainly because that the facts that they were at incorrect locations were nothing to do with the main goal of their dialogues. This was done by followers even though they were well aware of there were time delays in their mutual beliefs. In terms of the accommodation rule, execution postponement, silent adjustment, and unconfirmed might have been chosen to be done to increase accuracies, effectivenesses, and/or effeciencies. Unconfirmed, which is a DS that lacks any sign of completing tasks, may also be explained by the accommodation rule. By beginning a next DS and not hearing any sign of trouble, givers may be able to assume that his or her follower is in a right track. Likewise, hearing a next DS and not giving any sign of trouble allows a follower to think that the giver knows that he or she is in a right track.

- (1) 01:08:752-01:12:416 G: eto Isi-no-sabaku-no ima hidarisita-no-hasi desi-ta-ne You are at the lower left end of Stone desert, right?
- (2) 01:12:399-01:13:151 F: hai sita-ni i-masu Yes, I am at the bottom.
- (3) 01:13:210-01:14:130 F: (wrote route) [execution]

Fig. 14. Excerpt of silent adjustment



Fig. 15. Example of silent adjustment

A. Necessity of a model of understanding utterance

Although JMTDC may not represent typical situations, if conversational behaviors and underlying inference processes in JMTDC were applicable to general conversations, it is valuable to model the behavioral and inference processes found in the present study. There were delays in time between the process of constructing shared beliefs made by utterances and drawing routes in all of delayed acceptance, premature sign of completion, execution postponement, silent adjustment, unconfirmed, and indirection found in the present study. If routes draws by followers represents understandings and interpretation of the utterance of follower, it seems necessary to model understanding of utterance and constructing (superficial) mutual beliefs independently. In order to develop a model

that mimics how people conversationally and behaviourally acts in collaborative tasks, it is necessary to construct separate models for understanding utterances and constructing mutual beliefs and then effectively integrate these two models. The present research provide a novel view of understanding how mutual beliefs are constructed in collaborative tasks.

REFERENCES

- [1] H. P. Grice, "Meaning", The philosophical review, 66 (3), 1957, pp.377-388.
- [2] J. L. Austin, How to do things with words. Oxford university press, 1975.
- [3] J. R. Searle, Speech acts: An essay in the philosophy of language, Vol.626, Cambridge university press, 1969.
- [4] Y. Horiuchi, Y. Nakano, H. Koiso, M. Ishizaki, H. Suzuki, M. Okada, M. Naka, S. Tutiya, and A. Ichikawa, "The design and statistical characterization of the Japanese map task dialogue corpus," Journal of Japanese Society for Artificial Intelligence, vol. 14, no. 2, 1999, pp.261 272.
- [5] A. H. Anderson, et al. "The HCRC map task corpus", Language and speech 34 (4), 1991, pp.351-366.
- B. J. Grosz, and C. L. Sidner, "Attention, Intentions, and the Structure of [6] Discourse", Computational Linguistics, Vol.12, No.3, 1986, pp.175-204.
- [7] J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, A. Anderson, "HCRC Dialog Structure Coding Manual", Universities of Edinburgh and Glasgow, 1996.
- D. R. Traum, "A computational theory of grounding in natural language conversation". Ph. D. thesis, University of Rochester, 1994. [8]
- [9] J. F. Allen, and C. R. Perrault, "Analyzing Intention in Utterances", Artificial Intelligence, 15 (3), 1980, pp.143-178.
- [10] P. R. Cohen, C. R. Perrault, "Elements of a Plan–Based Theory of Speech Acts", Cognitive Science, 3, 1979, pp.177–212. [11] P. R. Cohen, H. J. Levesque, "Confirmations and Joint Action", In
- Proceedings IJCAI, 91, 1991, pp.951-957.
- [12] H. H. Clark, Using language, Cambridge University Press, 1996.
- [13] H. Lausberg, and H. Sloetjes, "Coding gestural behavior with the NEUROGES-ELAN system", Behavior Research Methods, Instruments, & Computers, 41 (3), 2009, pp.841-849. doi:10.3758/BRM.41.3.591.
- [14] D. Lewis, "Scorekeeping in a language game", Journal of Philosophical Logic, 8 (1), 1979, pp.339-359.