

# Identifying Repair Targets in Action Control Dialogue

Kotaro Funakoshi and Takenobu Tokunaga

Department of Computer Science,

Tokyo Institute of Technology

2-12-1 Oookayama Meguro, Tokyo, JAPAN

{koh,take}@cl.cs.titech.ac.jp

## Abstract

This paper proposes a method for dealing with repairs in action control dialogue to resolve participants' misunderstanding. The proposed method identifies the repair target based on common grounding rather than surface expressions. We extend Traum's grounding act model by introducing degree of groundedness, and partial and mid-discourse unit grounding. This paper contributes to achieving more natural human-machine dialogue and instantaneous and flexible control of agents.

## 1 Introduction

In natural language dialogue, misunderstanding and its resolution is inevitable for the natural course of dialogue. The past research dealing with misunderstanding has been focused on the dialogue involving only utterances. In this paper, we discuss misunderstanding problem in the dialogue involving participant's actions as well as utterances. In particular, we focus on misunderstanding in *action control dialogue*.

Action control dialogue is a kind of task-oriented dialogue in which a *commander* controls the actions<sup>1</sup> of other agents called *followers* through verbal interaction.

This paper deals with *disagreement repair initiation utterances*<sup>2</sup> (DRIUs) which are used by commanders to resolve followers' misunderstandings<sup>3</sup>, or to correct commanders' previous erroneous utterances. These are so called third-turn

<sup>1</sup>We use the term "action" for the physical behavior of agents except for speaking.

<sup>2</sup>This denomination is lengthy and may be still controversial. However we think this is most descriptively adequate for the moment.

<sup>3</sup>Misunderstanding is a state where miscommunication has occurred but participants are not aware of this, at least initially (Hirst et al., 1994).

repair (Schegloff, 1992). Unlike in ordinary dialogue consisting of only utterances, in action control dialogue, followers' misunderstanding could be manifested as their inappropriate actions in response to a given command.

Let us look at a sample dialogue (1.1–1.3). Utterance (1.3) is a DRIU for repairing V's misunderstanding of command (1.1) which is manifested by his action performed after saying "OK" in (1.2).

(1.1) U: Put the red book on the shelf to the right.

(1.2) V: OK. <V performs the action>

(1.3) U: Not that.

It is not easy for machine agents to understand DRIUs because they can sometimes be so elliptical and context-dependent that it is difficult to apply traditional interpretation methodology to DRIUs.

In the rest of this paper, we describe the difficulty of understanding DRIUs and propose a method to identify repair targets. The identification of repair targets plays a key role in understanding DRIUs and this paper is intensively focused on this issue.

## 2 Difficulty of Understanding DRIUs

Understanding a DRIU consists of repair target identification and repair content interpretation. Repair target identification identifies a target to be repaired by the speaker's utterance. Repair content interpretation recovers the speaker's intention by replacing the identified repair target with the correct one.

One of the major source of difficulties in understanding DRIUs is that they are often elliptical. Repair content interpretation depends heavily on repair targets but the information to identify repair targets is not always mentioned explicitly in DRIUs.

Let us look at dialogue (1.1–1.3) again. The DRIU (1.3) indicates that V failed to identify U’s intended object in utterance (1.1). However, (1.3) does not explicitly mention the repair target, i.e., either *book* or *shelf* in this case.

The interpretation of (1.3) changes depending on when it is uttered. More specifically, the interpretation depends on the local context and the situation when the DRIU is uttered. If (1.3) is uttered when V is reaching for a book, it would be natural to consider that (1.3) is aimed at repairing V’s interpretation of “the book”. On the other hand, if (1.3) is uttered when V is putting the book on a shelf, it would be natural to consider that (1.3) is aimed at repairing V’s interpretation of “the shelf to the right”.

Assume that U uttered (1.3) when V was putting a book in his hand on a shelf, how can V identify the repair target as *shelf* instead of *book*? This paper explains this problem on the basis of common grounding (Traum, 1994; Clark, 1996). Common grounding or shortly grounding is the process of building mutual belief among a speaker and hearers through dialogue. Note that in action control dialogue, we need to take into account not only utterances but also followers’ actions. To identify repair targets, we keep track of states of grounding by treating followers’ actions as grounding acts (see Section 3). Suppose V is placing a book in his hand on a shelf. At this moment, V’s interpretation of “the book” in (1.1) has been already grounded, since U did not utter any DRIU when V was taking the book. This leads to the interpretation that the repair target of (1.1) is *shelf* rather than already grounded *book*.

### 3 Grounding

This section briefly reviews the grounding acts model (Traum, 1994) which we adopted in our framework. We will extend the grounding act model by introducing degree of groundedness that have a quaternary distinction instead of the original binary distinction. The notions of partial grounding and mid-discourse unit grounding are also introduced for dealing with action control dialogue.

#### 3.1 Grounding Acts Model

The grounding acts model is a finite state transition model to dynamically compute the state of grounding in a dialogue from the viewpoint of

each participant.

This theory models the process of grounding with a theoretical construct, namely the *discourse unit* (DU). A DU is a sequence of *utterance units* (UUs) assigned *grounding acts* (GAs). Each UU in a dialogue has at least one GA, except fillers or several cue phrases, which are considered useful for turn taking but not for grounding. Each DU has an *initiator* (I) who opened it, and other participants of that DU are called *responders* (R).

Each DU is in one of seven states listed in Table 1 at a time. Given one of GAs shown in Table 2 as an input, the state of DU changes according to the current state and the input. A DU starts with a transition from initial state S to state 1, and finishes at state F or D. DUs in state F are regarded as grounded.

Analysis of the grounding process for a sample dialogue is illustrated in Figure 1. Speaker B can not understand the first utterance by speaker A and requests a repair (ReqRep-R) with his utterance. Responding to this request, A makes a repair (Repair-I). Finally, B acknowledges to show he has understood the first utterance and the discourse unit reaches the final state, i.e., state F.

State	Description
S	Initial state
1	Ongoing
2	Requested a repair by a responder
3	Repaired by a responder
4	Requested a repair by the initiator
F	Finished
D	Canceled

Table 1: DU states

Grounding act	Description
Initiate	Begin a new DU
Continue	Add related content
Ack	Present evidences of understanding
Repair	Correct misunderstanding
ReqRepair	Request a repair act
ReqAck	Request an acknowledge act
Cancel	Abandon the DU

Table 2: Grounding acts

UU	DUI	
A : Can I speak to Jim Johnstone please?	Init-I	1
B : Senior?	ReqRep-R	2
A : Yes	Repair-I	1
B : Yes	Ack-R	F

Figure 1: An example of grounding (Ishizaki and Den, 2001)

### 3.2 Degree of Groundedness and Evidence Intensity

As Traum admitted, the binary distinction between grounded and ungrounded in the grounding acts model is an *oversimplification* (Traum, 1999). Repair target identification requires more finely defined degree of groundedness. The reason for this will be elucidated in Section 5.

Here, we will define the four levels of evidence intensity and equate these with degrees of groundedness, i.e., if an utterance is grounded with evidence of level  $N$  intensity, the degree of groundedness of the utterance is regarded as level  $N$ .

#### (2) Levels of evidence intensity

**Level 0:** No evidence (i.e., not grounded).

**Level 1:** The evidence shows that the responder thinks he understood the utterance. However, it does not necessarily mean that the responder understood it correctly. E.g., the acknowledgment “OK” in response to the request “turn to the right.”

**Level 2:** The evidence shows that the responder (partially) succeeded in transferring surface level information. It does not yet ensure that the interpretation of the surface information is correct. E.g., the repetition “to the right” in response to the request “turn to the right.”

**Level 3:** The evidence shows that the responder succeeded in interpretation. E.g., turning to the right as the speaker intended in response to the request “turn to the right.”

### 3.3 Partial and mid-DU Grounding

In Traum’s grounding model, the content of a DU is uniformly grounded. However, things in the same DU should be more finely grounded at various levels individually. For example, if one acknowledged by saying “to the right” in response to the command “put the red chair to the right of the table”, *to\_the\_right\_of* should be regarded as grounded at Level 2 even though other parts of the request are grounded at Level 1.

In addition, in Traum’s model, the content of a DU is grounded all at once when the DU reaches the final state, F. However, some elements in a DU can be grounded even though the DU has not yet reached state F. For example, if one requested a

repair as “to the right of what?” in response to the command “put the red chair to the right of the table”, *to\_the\_right\_of* should be regarded as grounded at level 2 even though *table* has not yet been grounded.

Although Traum admitted these problems existed in his model, he retained it for the sake of simplicity. However, such partial and mid-DU grounding is necessary to identify repair targets. We will describe the usage of these devices to identify repair targets in Section 5. In brief, when a level 3 evidence is presented by the follower and negative feedback (i.e., DRIUs) is not provided by the commander, only propositions supported by the evidence are considered to be grounded even though the DU has not yet reached state F.

## 4 Treatment of Actions in Dialogue

In general, past work on discourse has targeted dialogue consisting of only utterances, or has considered actions as subsidiary elements. In contrast, this paper targets action control dialogue, where actions are considered to be primary elements of dialogue as well as utterances.

Two issues have to be mentioned for handling action control dialogue in the conventional sequential representation as in Figure 1. We will introduce assumptions (3) and (4) as shown below.

### Overlap between utterances and actions

Actions in dialogue do not generally obey turn allocation rules as Clark pointed out (Clark, 1996). In human-human action control dialogue, followers often start actions in the middle of a commander’s utterance. This makes it difficult to analyze discourse in sequential representation. Given this fact, we impose the three assumptions on followers as shown in (3) so that followers’ actions will not overlap the utterances of commanders. These requirements are not unreasonable as long as followers are machine agents.

#### (3) Assumptions on follower’s actions

- (a) The follower will not commence action until turn taking is allowed.
- (b) The follower immediately stops the action when the commander interrupts him.
- (c) The follower will not make action as primary elements while speaking.<sup>4</sup>

<sup>4</sup>We regard gestures such as pointing as secondary ele-

## Hierarchy of actions

An action can be composed of several sub-actions, thus has a hierarchical structure. For example, making tea is composed of boiling the water, preparing the tea pot, putting tea leaves in the pot, and pouring the boiled water into it, and so on. To analyze actions in dialogue as well as utterances in the traditional way, a unit of analysis should be determined. We assume that there is a certain granularity of action that human can recognize as primitive. These actions would correspond to basic verbs common to humans such as “walk”, “grasp”, “look”, etc. We call these actions *fundamental actions* and consider them as UUs in action control dialogue.

### (4) Assumptions on fundamental actions

In the hierarchy of actions, there is a certain level consisting of fundamental actions that human can commonly recognize as primitives. Fundamental actions can be treated as units of primary presentations in an analogy with utterance units .

## 5 Repair Target Identification

In this section, we will discuss how to identify the repair target of a DRIU based on the notion of grounding. The following discussion is from the viewpoint of the follower.

Let us look at a sample dialogue (5.1–5.5), where U is the commander and V is the follower. The annotation  $Ack_1-R:F$  in (5.2) means that (5.2) has grounding act  $Ack$  by the responder (R) for DU1 and the grounding act made DU1 enter state F. The angle bracketed descriptions in (5.3) and (5.4) indicate the fundamental actions by V.

Note that thanks to assumption (4) in Section 4, a fundamental action itself can be considered as a UU even though the action is performed without any utterances.

(5.1) U: Put the red ball on the left box. ( $Init_1-I:1$ )

(5.2) V: Sure. ( $Ack_1-R:F$ )

(5.3) V: <V grasps the ball> ( $Init_2-I:1$ )

(5.4) V: <V moves the ball> ( $Cont_2-I:1$ )

(5.5) U: Not that. ( $Repair_1-R:3$ )

The semantic content of (5.1) can be represented as a set of propositions as shown in (6).

When they are presented in parallel with speech. Therefore, this constraint does not apply to them.

(6)  $\alpha = Request(U, V, Put(\#Agt1, \#Obj1, \#Dst1))$

(a)  $speechActType(\alpha) = Request$

(b)  $presenter(\alpha) = U$

(c)  $addressee(\alpha) = V$

(d)  $actionType(content(\alpha)) = Put$

(e)  $agent(content(\alpha)) = \#Agt1$ ,  
 $referent(\#Agt1) = V$

(f)  $object(content(\alpha)) = \#Obj1$ ,  
 $referent(\#Obj1) = Ball1$

(g)  $destination(content(\alpha)) = \#Dst1$ ,  
 $referent(\#Dst1) = Box1$

$\alpha$  represents the entire content of (5.1). Symbols beginning with a lower case letter are function symbols. For example, (6a) means the speech act type for  $\alpha$  is “Request”. Symbols beginning with an upper case letter are constants. “Request” is the name of a speech act type and “Move” is that of fundamental action respectively. U and V represents dialogue participants and “Ball1” represents an entity in the world. Symbols beginning with # are notional entities introduced in the discourse and are called discourse referents. A discourse referent represents something referred to linguistically. During a dialogue, we need to connect discourse referents to entities in the world, but in the middle of the dialogue, some discourse referents might be left unconnected. As a result we can talk about entities that we do not know. However, when one takes some actions on a discourse referent, he must identify the entity in the world (e.g., an object or a location) corresponding to the discourse referent. Many problems in action control dialogue are caused by misidentifying entities in the world.

Follower V interprets (5.1) to obtain (6), and prepares an action plan (7) to achieve “Put( $\#Agt1, \#Obj1, \#Dst1$ )”. Plan (7) is executed downward from the top.

(7) Plan for Put( $\#Agt1, \#Obj1, \#Dst1$ )

Grasp( $\#Agt1, \#Obj1$ ),

Move( $\#Agt1, \#Obj1, \#Dst1$ ),

Release( $\#Agt1, \#Obj1$ )

Here, (5.1–5.5) are reformulated as in (8.1–8.5). “Perform” represents performing the action.

(8.1) U: Request(U, V, Put( $\#Agt1, \#Obj1, \#Dst1$ ))

(8.2) V: Accept(V, U,  $\alpha$ )

(8.3) V: Perform(V, U, Grasp( $\#Agt1, \#Obj1$ ))

(8.4) V: Perform(V, U, Move(#Agt1, #Obj1, #Dst1))

(8.5) U: Inform(U, V, incorrect(X))

To understand DRIU (5.5), i.e., (8.5), follower V has to identify repair target X in (8.5) referred to as “that” in (5.5). In this case, the repair target of (5.5) X is “the left box”, i.e., #Dst1.<sup>5</sup> However, the pronoun “that” cannot be resolved by anaphora resolution only using textual information.

We treat propositions, or bindings of variables and values, such as (6a–6g), as the minimum granularity of grounding because the identification of repair targets requires that granularity. We then make the following assumptions concerning repair target identification.

(9) Assumptions on repair target identification

- (a) Locality of elliptical DRIUs: The target of an elliptical DRIU that interrupted the follower’s action is a proposition that is given an evidence of understanding by the interrupted action.
- (b) Instancy of error detection: A dialogue participant observes his dialogue constantly and actions presenting strong evidence (Level 3). Thus, when there is an error, the commander detects it immediately once an action related to that error occurs.
- (c) Instancy of repairs: If an error is found, the commander immediately interrupts the dialogue and initiates a repair against it.
- (d) Lack of negative evidence as positive evidence: The follower can determine that his interpretation is correct if the commander does not initiate a repair against the follower’s action related to the interpretation.
- (e) Priority of repair targets: If there are several possible repair targets, the least grounded one is chosen.

(9a) assumes that a DRIU can only be elliptical when it presupposes the use of local context to identify its target. It also predicts that if the target of a repair is neither local nor accessible within local information, the DRIU will not be elliptical depending on local context but contain explicit and

<sup>5</sup>We assume that there is a sufficiently long interval between the initiations of (5.4) and (5.5).

sufficient information to identify the target. (9b) and (9c) enable (9a).

Nakano *et al.* (2003) experimentally confirmed that we observe negative responses as well as positive responses in the process of grounding. According to their observations, speakers continue dialogues if negative responses are not found even when positive responses are not found. This evidence supports (9d).

An intuitive rationale for (9e) is that an issue with less proof would more probably be wrong than one with more proof.

Now let us go through (8.2) to (8.5) again according to the assumptions in (9). First,  $\alpha$  is grounded at intensity level 1 by (8.2). Second, V executes Grasp(#Agt1, #Obj1) at (8.3). Because V does not observe any negative response from U even after this action is completed, V considers that the interpretations of #Agt1 and #Obj1 have been confirmed and grounded at intensity level 3 according to (9d) (this is the partial and mid-DU grounding mentioned in Section 3.3). After initiating Move(#Agt1, #Obj1, #Dst1), V is interrupted by commander U with (8.5) in the middle of the action.

V interprets elliptical DRIU (5.5) as “Inform(S, T, incorrect(X))”, but he cannot identify repair target X. He tries to identify this from the discourse state or context. According to (9a), V assumes that the repair target is a proposition that its interpretation is demonstrated by interrupted action (8.4). Due to the nature of the word “that”, V knows that possible candidates are not types of action or the speech act but discourse referents #Agt1, #Obj1 and #Dst1<sup>6</sup>. Here, #Agt1 and #Obj1 have been grounded at intensity level 3 by the completion of (8.3). Now, (9e) tells V that the repair target is #Dst1, which has only been grounded at intensity level 1<sup>7</sup>.

(10) below summarizes the method of repair target identification based on the assumptions in (9).

(10) Repair target identification

<sup>6</sup>We have consistently assumed Japanese dialogues in this paper although examples have been translated into English. “That” is originally the pronoun “*sotti*” in Japanese, which can only refer to objects, locations, or directions, but cannot refer to actions.

<sup>7</sup>There are two propositions concerned with #Dst1: destination(content( $\alpha$ )) = #Dst1 and referent(#Dst1) = Box1. However if dest(content( $\alpha$ )) = #Dst1 is not correct, this means that V grammatically misinterpreted (8.1). It seems hard to imagine for participants speaking in their mother tongue and thus one can exclude dest(content( $\alpha$ )) = #Dst1 from the candidates of the repair target.

- (a) Specify the possible types of the repair target from the linguistic expression.
- (b) List the candidates matching the types determined in (10a) from the latest presented content.
- (c) Rank candidates based on groundedness according to (9e) and choose the top ranking one.

### Dependencies between Parameters

The follower prepares an action plan to achieve the commander’s command as in plan (7). Here, the planned actions can contain parameters not directly corresponding to the propositions given by the commander. Sometimes a selected parameter by using (10) is not the true target but the dependent of the target. Agents must retrieve the true target by recognizing dependencies of parameters.

For example, assume a situation where objects are not within the follower’s reach as shown in Figure 2. Then, the commander issues command (6) to the follower (Agent1 in Figure 2) and he prepares an action plan (11).

(11) Agent1’s plan (partial) for (6) in Figure 2.

```

Walk(#Agt1, #Dst1),
Grasp(#Agt1, #Obj1),
...

```

The first Walk is a prerequisite action for Grasp and #Dst1 depends on #Obj1. In this case, if referent(#Obj1) is Object1 then referent(#Dst1) is Position1, or if referent(#Obj1) is Object2 then referent(#Dst1) is Position2. Now, assume that the commander intends referent(#Obj1) to be Object2 with (6), but the follower interprets this as referent(#Obj1) = Object1 (i.e., referent(#Dst1) = Position1) and performs Walk(#Agt1, #Dst1). The commander then observes the follower moving toward a direction different from his expectation and infers the follower has misunderstood the target object. He, then, interrupts the follower with the utterance “not that” at the timing illustrated in Figure 3. Because (10c) chooses #Dst2 as the repair target, the follower must be aware of the dependencies between parameters #Dst1 and #Obj1 to notice his misidentification of #Obj1.

## 6 Implementation and Some Problems

We implemented the repair target identification method described in Section 5 into our prototype

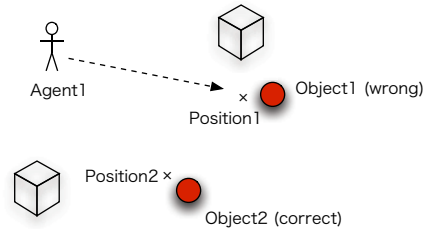


Figure 2: Situation with dependent parameters

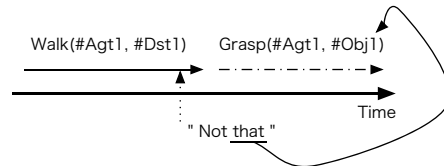


Figure 3: Dependency between parameters

dialogue system (Figure 4). The dialogue system has animated humanoid agents in its visualized 3D virtual world. Users can command the agent by speech to move around and relocate objects.



Figure 4: Snapshot of the dialogue system

Because our domain is rather small, current possible repair targets are agents, objects and goals of actions. According to the qualitative evaluation of the system through interaction with several subjects, most of the repair targets were correctly identified by the proposed method described in Section 5. However, through the evaluation, we found several important problems to be solved as below.

### 6.1 Feedback Delay

In a dialogue where participants are paying attention to each other, the lack of negative feedback can be considered as positive evidence (see (9d)). However, it is not clear how long the system needs to wait to consider the lack of negative feedback as positive evidence. In some cases, it will be not appropriate to consider the lack of negative feedback

as positive evidence immediately after an action has been completed. Non-linguistic information such as nodding and gazing should be taken into consideration to resolve this problem as (Nakano et al., 2003) proposed.

Positive feedback is also affected by delay. When one receives feedback shortly after an action is completed and begins the next action, it may be difficult to determine whether the feedback is directed to the completed action or to the just started action.

## 6.2 Visibility of Actions

The visibility of followers' actions must be considered. If the commander cannot observe the follower's action due to environmental conditions, the lack of negative feedback cannot be positive evidence for grounding.

For example, assume the command "bring me a big red cup from the next room" is given and assume that the commander cannot see the inside of the next room. Because the follower's fundamental action of taking a cup in the next room is invisible to the commander, it cannot be grounded at that time. They have to wait for the return of the follower with a cup.

## 6.3 Time-dependency of Grounding

Utterances are generally regarded as points on the time-line in dialogue processing. However, this approximation cannot be applied to actions. One action can present evidences for multiple propositions but it will present these evidences at considerably different time. This affects repair target identification.

Let us look at an action `Walk(#Agt, #Dst)`, where agent `#Agt` walks to destination `#Dst`. This action will present evidence for "who is the intended agent (`#Agt`)" at the beginning. However, the evidence for "where is the intended position (`#Dst`)" will require the action to be completed. However, if the position intended by the follower is in a completely different direction from the one intended by the commander, his misunderstanding will be evident at a fairly early stage of the action.

## 6.4 Differences in Evidence Intensities between Actions

Evidence intensities vary depending on the characteristics of actions. Although the symbolic description of actions such as (12) and (13) does not explicitly represent differences in intensity, there

is a significant difference between (12) where `#Agent` looks at `#Object` at a distance, and (13) where `#Agent` directly contacts `#Object`. Agents must recognize these differences to conform with human recognition and share the same state of grounding with participants.

(12) `LookAt(#Agent, #Object)`

(13) `Grasp(#Agent, #Object)`

## 6.5 Other Factors of Confidence in Understanding

Performing action can provide strong evidence of understanding and such evidence enables participants to have strong confidence in understanding. However, other factors such as linguistic constraints (not limited to surface information) and plan/goal inference can provide confidence in understanding without grounding. Such factors of confidence also must be incorporated to explain some repairs.

Let us see a sample dialogue below, and assume that follower V missed the word *red* in (14.3).

(14.1) U: Get the white ball in front of the table.

(14.2) V: OK. <V takes a white ball>

(14.3) U: Put it on the (red) table.

(14.4) V: Sure. <V puts the white ball holding in his hand on a *non-red* table>

(14.5) U: I said red.

When commander U repairs V's misunderstanding by (14.5), V cannot correctly decide that the repair target is not "it" but "the (red) table" in (14.3) by using the proposed method, because the referent of "it" had already been in V's hand and no explicit action choosing a ball was performed after (14.3). However, in such a situation we seem to readily doubt misunderstanding of "the table" because of strong confidence in understanding of "it" that comes from outside of grounding process. Hence, we need a unified model of confidence in understanding that can map different sources of confidence into one dimension. Such a model is also useful for clarification management of dialogue systems.

## 7 Discussion

### 7.1 Advantage of Proposed Method

The method of repair target identification proposed in this paper less relies on surface information to identify targets. This is advantageous

against some sort of misrecognitions by automatic speech recognizers and contributes to the robustness of spoken dialogue systems.

Only surface information is generally insufficient to identify repair targets. For example, assume that there is an agent acting in response to (15) and his commander interrupts him with (16).

(15) Put the red ball on the table

(16) Sorry, I meant blue

If one tries to identify the repair target with surface information, the most likely candidate will be “the red ball” because of the lexical similarity. Such methods easily break down. They cannot deal with (16) after (17). If, however, one pays attention to the state of grounding as our proposed method, he can decide which one is likely to be repaired “the red ball” or “the green table” depending on the timing of the DRIU.

(17) Put the red ball on the green table

## 7.2 Related Work

McRoy and Hirst (1995) addressed the detection and resolution of misunderstandings on speech acts using abduction. Their model only dealt with speech acts and did not achieve our goals.

Ardissono *et al.* (1998) also addressed the same problem but with a different approach. Their model could also handle misunderstanding regarding domain level actions. However, we think that their model using coherence to detect and resolve misunderstandings cannot handle DRIUs such as (8.5), since both possible repairs for #Obj1 and #Dst1 have the same degree of coherence in their model.

Although we did not adopt this, the notion of QUD (questions under discussion) proposed by Ginzburg (Ginzburg, 1996) would be another possible approach to explaining the problems addressed in this paper. It is not yet clear whether QUD would be better or not.

## 8 Conclusion

Identifying repair targets is a prerequisite to understand disagreement repair initiation utterances (DRIUs). This paper proposed a method to identify the target of a DRIU for conversational agents in action control dialogue. We explained how a repair target is identified by using the notion of common grounding. The proposed method has been

implemented in our prototype system and evaluated qualitatively. We described the problems found in the evaluation and looked at the future directions to solve these problems.

## Acknowledgment

This work was supported in part by the Ministry of Education, Science, Sports and Culture of Japan as the Grant-in-Aid for Creative Basic Research No. 13NP0301.

## References

- L. Ardissono, G. Boella, and R. Damiano. 1998. A plan based model of misunderstandings in cooperative dialogue. *International Journal of Human-Computer Studies*, 48:649–679.
- Herbert H. Clark. 1996. *Using Language*. Cambridge University Press.
- Jonathan Ginzburg. 1996. Interrogatives: questions, facts and dialogue. In Shalom Lappin, editor, *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.
- G. Hirst, S. McRoy, P. Heeman, P. Edmonds, and D. Horton. 1994. Repairing conversational misunderstandings and non-understandings. *Speech Communication*, 15:213–230.
- Masato Ishizaki and Yasuharu Den. 2001. *Danwa to taiwa (Discourse and Dialogue)*. University of Tokyo Press. (In Japanese).
- Susan Weber McRoy and Graeme Hirst. 1995. The repair of speech act misunderstandings by abductive inference. *Computational Linguistics*, 21(4):435–478.
- Yukiko Nakano, Gabe Reinstein, Tom Stocky, and Justine Cassell. 2003. Towards a model of face-to-face grounding. In Erhard Hinrichs and Dan Roth, editors, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 553–561.
- E. A. Schegloff. 1992. Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5):1295–1345.
- David R. Traum. 1994. *Toward a Computational Theory of Grounding*. Ph.D. thesis, University of Rochester.
- David R. Traum. 1999. Computational models of grounding in collaborative systems. In *Working Papers of AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, pages 137–140.