Task Context Dependency of Explanation Strategy in Instruction Dialogue

Yukiko, I. Nakano and Tsuneaki Kato
NTT Information and Communication Systems Labs.
1-2356 Take Yokosuka-shi Kanagawa 238-03, Japan
{yukiko, kato}@nttly.isl.ntt.jp

Abstract

When an expert explains some operation to a novice, different explanation strategies are used for first time explanation, review explanation, and re-explanation for recovery from communication failure. In this paper, first by analyzing dialogue instruction dialogue corpus, we reveal how task context as well as discourse context determines the experts’ explanation strategy, then, based on the empirical results, we also introduce a mechanism for selecting the most appropriate utterance content and dialogue control strategy for an instruction dialogue.

Introduction

In communication conducted in the real world, conversation participants monitor the partner’s understanding based on not only the linguistic response of the partner, but also the partner’s behavior or the situation resulting from the behavior. Therefore, it is supposed that the linguistic production in conversation is affected by the context defined by the expert’s linguistic behavior and physical state of the task. For example, when two agents work together to operate some physical object, the linguistic communication would be affected by whether the operation succeeds or fails. In particular, in expert-novice explanation dialogues in which an expert agent explains the operation of some physical object to a novice agent, the expert’s explanations depend on whether the novice succeeds or fails. The expert re-explains the operation and guides the novice to repeat the operation if he recognizes that the novice failed to understand the operation correctly. The re-explanation after failure, would differ from the original explanation even though the expert is explaining the same operation. If the novice is successful in the first trial, the explanation for reviewing the procedure also differs from the original explanation and the re-explanation as well. Therefore, it is considered that the expert modifies the explanation according to the observable success or failure of the novice’s operation.

Consider the following dialogue between an expert and a novice about setting up an answering machine. These were originally performed in Japanese.

(1): First-time explanation (Practice Step)
E: Pick up the receiver.
N: Yes.
E: Release the response button.
N: Yes.
E: Then the response button should start blinking.
N: Yes.
E: Speak the answering message.

(2): Second-time explanation after a successful attempt (Review Step)
E: Pick up the receiver, release the response button, and speak the answering message.
N: Yes.

In dialogue (1), the expert explains the procedure step by step with additional information that helps the novice’s understanding. This strategy makes it possible to explain while confirming the novice’s understanding. On the other hand, in dialogue (2), the explanation is simpler. The direction utterance has no additional information, and the three actions are mentioned in one turn. Considering the above discussion, the explanatory dialogue system should be able to modify the explanation content and dialogue strategy according to discourse context (dialogue history) and task context such as re-explanation or review.

There are some works concerned with modifying explanatory dialogues according to context. As for the selection of utterance content, (Moore 1995) proposed some heuristics to select contextually appropriate additional information that give some background or elaboration to achieve the communicative goals for generating explanation texts. She proposed that if the communicative goal fails, some recovery heuristics are used, such as giving some example or analogy as additional information and expanding any additional information not yet mentioned. She, however, was not concerned with the issue of dialogue organization. On the other hand, (Cawsey 1990;
Table 1: The characteristics of task context

<table>
<thead>
<tr>
<th></th>
<th>Task history weight</th>
<th>Understanding level weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Review</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Re-explanation</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Cawsey (1993) addressed the issue of dialogue management. She discussed the ideal style of human-computer interaction. Her dialogue system requires the user's acknowledgment after each utterance (typically a sentence in length). It remains a question, however, how much information should be provided in one turn and when the turn should be changed.

Therefore, in order to construct instruction dialogue system, it is important to clarify the following two points:

- how the amount and the kind of additional information depends on the context,
- how the dialogue is organized, especially when turns should be released.

This study analyzes a dialogue corpus to reveal how the expert's explanation strategy depends on the task context and the discourse context. Then, based on the result of the corpus analysis, we propose a mechanism for selecting the explanation utterances of an explanatory dialogue according to the observable context.

Framework
Factors that Determine the Characteristics of Dialogue

We consider task context and discourse context as the factors determining the expert's language performance such as utterance content and the dialogue organization strategy.

(a) Task context: Task context is the state of task execution and is classified into three categories;
- Practice step: Original explanation along with coaching the novice.
- Review step: If the novice succeeds in performing the operation, the expert makes the novice try it again for review purposes.
- Re-explanation step: If the novice fails to perform the operation, the expert re-explains the operation.

These task context categories can be represented by the combination of two parameters: task history and the understanding level of the novice.

Task History: How many times the task has been explained or executed.

Understanding level of the novice: How well the novice understands task execution.

These two parameters are used to represent task context as shown in Table 1. The practice step is the first time explanation and the novice is not expected to understand the task procedure, therefore this step can be represented as the combination of negative task history weight and negative understanding level weight. On the other hand, the review step has positive weight for both parameters. The re-explanation step has positive task history weight and negative understanding level weight.

(b) Discourse context: Discourse context is the history of the dialogue. We characterize the utterance content with regard to whether it was already mentioned or not by referring to the discourse context:

- Redundant information: Utterance content that has already been mentioned. In terms of pragmatics, it is informationally redundant (Grice 1975).
- Novel information: Utterance content not previously mentioned.

Aspects of Corpus Analysis

We established the framework for analyzing the utterance content and dialogue strategy performed by experts.

Utterance Content of Explanation (Moore 1989; Moore 1995) treats explanation content as plan operators and describes them based on RST (rhetorical structure theory) (Mann 1987) which characterizes text structure in terms of the relation that holds between parts of a text. She realized these relations in a plan operator. In the plan operator, the main topic to be expressed is defined as the nucleus and the additional information needed to achieve the communicative goal are defined as satellites. We incorporate this framework, and investigate what and how many satellites are expanded as additional information according to task context and discourse context. We define the following categories of rhetorical relations based on the RST (Rhetorical Structure Theory) (Mann 1987; Hovy 1993).

(a) Cause/Result: Referring to the result of the action and its meaning, or referring to the situation causing the action.
(b) **Elaboration**: Elaboration of the action and the characteristics or examples of objects.

(c) **Condition**: Referring to the condition required to perform the action.

(d) **Sequence**: Referring to the sequence of actions or events.

(e) **Others**: Relations besides those types mentioned above, including circumstance relation as used in (Mann 1987; Hovy 1993).

**Dialogue Strategy** In the framework of discourse analysis (Stenström 1994), spoken interaction is described in terms of five hierarchical levels: transaction, exchange, turn, move, and act. Exchange is the smallest unit of interaction, and transaction consists of a sequence of exchanges. Turn is everything A says before B takes over and consists of the one or more moves that start or finish an exchange.

As one aspect of these dialogue strategies, we focus on the turn taking unit which is controlled by the expert. In this study, we regard a sentence as a turn taking unit in dialogue as does (Cawsey 1990). Therefore, the question is - does the amount of information provided in one sentence depend on context or not. We investigate how the satellites and the nucleus (action directions) are expressed. We set two expression forms; one is the turn-continue in which the utterance is continued to the next utterance content, the other is the turn-release in which the utterance is finished at this point and a chance for turn taking is given to the novice. In the following examples, pushing the response button is the action direction, and button blinking is the additional information that refers to the result of the action.

"After pushing the response button (turn-continue), it starts blinking (turn-release)."

"Push the response button (turn-release). Then it will start blinking (turn-release)."

**Empirical Study**

**Data Collection**

We collected explanation dialogues between experts and novices about installing an answering machine. Five experts participated in the data collection experiment. 56 novices received one explanation apiece. Therefore, 56 expert-novice dialogues were collected.

**Result of Empirical Study**

**Utterance Content Dependence on Context** The number of satellites provided was counted to examine how the amount of additional information depends on task context. The average number of additional explanations are shown in Figure 1. Additional information is added most frequently in the practice step (2.70), and least frequently in the review step (0.95). Therefore, it was found that the amount of additional information depends on task context ($F(2, 117)=19.344, p<0.0001$). We also examined the number of redundant additional and novel additional explanations in the review and re-explanation steps in order to examine the effect of the interaction between task context and discourse context. As shown in Figure 2, in the review step, redundant additional explanations...
are slightly more frequent than novel additional explanations (0.55 and 0.41 respectively), but the difference is not statistically significant. On the other hand, in the re-explanation step, novel additional explanations are more frequent than redundant additional explanations (1.25 and 0.55, respectively), and the difference is statistically significant \( t(38) = -2.252, p < 0.05 \).

Our empirical study found that, in the practice step, the expert tries to give sufficient additional information to allow the novice to succeed. While, in subsequent explanations, discourse context, in addition to task context, is taken into account in determining what satellites should be expanded. In the re-explanation step, novel information is preferred to redundant information. This result is consistent with the heuristic in (Moore 1995) which is that unexpanded additional information is used to recover communication failure. This result suggests that providing information not yet mentioned is more effective than repeating the same information when the user does not have complete understanding. These results indicate that the content of additional explanations is decided by the interaction between task context and discourse context.

**Dialogue Strategy**  In order to reveal that the turn taking unit depends on task context and discourse context, we examined how the utterance content is expressed. Figure 3 shows the percentage of turn-release in nucleus (action direction) in each task context. In the practice step, the percentage of turn-release is the highest (67% of the action directions were the turn-release type, that is, the utterance was terminated at that point). The turn-release frequency is lowest in the review step (54%) and it was 60% in the re-explanation step. The percentage of the turn-release (or turn-continue) for nucleus depended on the task context \( F(2,117)=3.073, p<0.003 \).

As for the expression form of satellites (additional information), result depended on the type of rhetorical relation used. “Condition” was generally expressed using turn-continue irrespective of task context or discourse context. It is assumed that the condition for a given action or event is conventionally expressed by using conditional sentences and this language usage convention overcomes the effect of context. As for the other types of additional information, however, expression preference depended on context. As shown in Figure 4, turn-release is most preferred in the practice step (80%), then re-explanation (62%), and least preferred in the review step (52%). The percentage of turn-continue, of course is the converse of this.

It is also expected that not only task context, but also discourse context affects the expression form of the utterance content. We investigated only “Cause/Result” because the corpus provided insufficient data as regards “Elaboration”. Figure 5 shows how the redundant and novel additional explanations about “Cause/Result” are expressed. In the re-

---

**Figure 3:** Percentage of turn-release in action directions

**Figure 4:** Percentage of turn-release in additional information expressions

**Figure 5:** Expression form of Cause/Result information
explanation step, novel additional explanations are more likely to be expressed as turn-release, not as turn-continue ($t(38) = 2.142, p < 0.05$).

From these results, it was found that in the practice step the expert prefers to provide both action directions and additional information in small chunks. However, in subsequent explanations, the information chunk size increases. This trend is strong in the review step.

**Discussion about Dialogue Redundancy**

In terms of pragmatics, if some piece of additional information has already been mentioned in the discourse, in other words, it is held in the discourse context, it is desirable not to generate it again in order to avoid redundancy (Grice 1975). The results of our empirical study, however, indicate that redundant information can be used. This gives a counterargument to the conversation maxim in (Grice 1975) and gives positive proof to (Walker 1992) who claims that informationally redundant utterances are necessary to establish mutual belief. Moreover, a remarkable point is that redundant information is not provided by repeating the same surface expression, but is presented together with other relevant statements and realized in several simple and shortened forms that are combined and expressed in a chunk.

**Applying the Empirical Results to Planning Instruction Dialogue**

The results of the empirical study reported above indicate that task context (represented by the combination of task history and novice's understanding level) and the discourse context (history of dialogue) determine the utterance content (especially additional information), and the dialogue organization strategy (especially turn taking unit), in instruction dialogues. Therefore, these three factors, task history, novice's understanding level, dialogue history, can be used as parameters to modify the instruction dialogue appropriately in instruction dialogue planning.

In order to plan an instruction dialogue, we separate the planning of utterance content from the planning of dialogue strategy because they are different aspects of language performance and we adopt different theories as the basis of analysis. One of the previous works on planning explanatory dialogue (Cawsey 1993) also separates content planning and dialogue planning and we basically follow Cawsey's approach. She uses domain dependent content plans and discourse analysis (DA) based dialogue plans. We adopt DA as the basis for dialogue planning, but we do not make the content plans domain dependent. Instead, we adopt RST as the theoretical basis of content plan in order to describe content plans more generally.

### Utterance Content Selection Rules

First, we discuss how to decide explanation content. The result of corpus analysis indicates that in the practice step, which has negative weight for both task-history and understanding level (Table 1), all candidates are novel, and they are most frequently expanded in this step. Novel information is less preferred in the review step which has negative weight for both task-history and understanding level. Redundant information is not preferred in any context. Therefore, the additional information selection preferences for novel information are defined as follows;

$$\text{Candidate Score} = \text{constant}(c1) + t_{h\text{-weight}} \cdot t_{h\text{-score}} + u_{l\text{-weight}} \cdot u_{l\text{-score}}$$

**Task history weight ($t_{h\text{-weight}}$):** The value related to how many times the task was explained and tried before.

**Understanding level weight ($u_{l\text{-weight}}$):**

The value related to how well the novice understands the procedure. It is assigned based on the execution result of the preceding session (success or failure).

**$t_{h\text{-score}}$ and $u_{l\text{-score}}$:** The values assigned to each candidate. Novel information is assigned negative score for both parameters.

For each candidate, if the sum of the weighted scores exceeds a threshold, the candidate is expanded. The polarity of each parameter and that of the composed vector are shown in Table 2. The polarity of candidate score is (+ +) for the practice step though it is (− −) for the review step. The ordering of candidate scores is consistent with the empirical result on the frequency of novel information explanation in Figure 1, Figure 2.

### Turn Taking Unit Decision Rules

We propose how to decide dialogue strategy. In empirical study, we focussed on how the turn taking unit is decided according to task context. It was found that the preference for turn-release decreased in the order Practice step, Re-explanation step, and then review step. As for continue-turn the preference order is the reverse of this. Therefore the preference rule can be described as follows (except for condition);

$$\text{Exchange (or Continue) Score} = \text{constant}(c1) + t_{h\text{-weight}} \cdot t_{h\text{-score}} + u_{l\text{-weight}} \cdot u_{l\text{-score}}$$

The $t_{h\text{-weight}}$ and the $u_{l\text{-weight}}$ are defined as before. Turn-release has a negative score in
Table 2: Preference for novel information explanation

<table>
<thead>
<tr>
<th>Novel information</th>
<th>t_h_score</th>
<th>uA_score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>t_h_weight</td>
<td>uA_weight</td>
</tr>
<tr>
<td>Review</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Re-explanation</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 3: Preferences for turn-release and turn-continue

<table>
<thead>
<tr>
<th>Turn-release</th>
<th>t_h_score</th>
<th>uA_score</th>
<th>Turn-continue</th>
<th>t_h_score</th>
<th>uA_score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice</td>
<td>t_h_weight</td>
<td>uA_weight</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Review</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Re-explanation</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Both $t_h$-score and $uA$-score because it is most preferred in novel context. Turn-continue has a positive score in both $t_h$-score and $uA$-score because it is most preferred in the review step. The plan operator (turn-release and turn-continue) that has the biggest weighted score is selected as the dialogue organization strategy. Table 3 shows the polarity for each parameter and that of the composed vector. In the practice step, the composed vector tends strongly to Turn-release. In the review step, the tendency is to Turn-continue. As for the Re-explanation step, the polarity depends on the weights. If the $t_h$-weight is very high, Turn-continue is preferred. On the other hand, if the $uA$-weight is very low then Turn-release is preferred. The baseline for expression form can be treated as a constant value ($cI$).

Example

We show an example of how an instruction dialogue is planned by using the proposed model. We incorporated the plan selection rules into a planning mechanism. We used a DA-based five level dialogue plan for dialogue planning and an RST-based content plan for content planning. The dialogue plan for dialogue (2) in the review step is shown in Figure 6. First, teaching-transaction is expanded, and then the teaching exchanges are expanded based on RST for deciding utterance content. The Cause/Result satellite is not expanded here (circled by dotted line) because it is omitted by the preference rule which states that use of redundant additional information is not preferred in the review step. After content planning is finished, the dialogue strategy is planned again. The plan is expanded from exchange level to the turn-taking level. In this dialogue, both the $t_h$-weight and $uA$-weight are high, so Request-turn-continue is selected based on the preference rule. In Request-turn-continue, acknowledgment by the user is not indispensable. Also in the next action direction, the Request-turn-continue is selected. Request-turn-release is selected for the third action direction because it is the last in the action sequence. At this time, the novice's acknowledgment is necessary.

Future Work

In this study, we used three context parameters in order to decide utterance content and the dialogue strategy. It is also necessary to consider other factors such as factors specific to utterance content. (Linden 1994; Linden 1995) proposes some ideas for deciding the grammatical form of some rhetorical relations by using the characteristics of action. In this paper, the issue of sentence boundary is treated using the clause combining strategy. Although Linden focuses on instruction text not instruction dialogue, there would be some discourse type (text or speech) independent and utterance content dependent factors that determine the utterance/sentence boundary. It is important to integrate Linden's proposals into our model.

The next point is about the limit of human memory. Regardless of how well the user understands the procedure, it may not be best to put the entire procedure explanation in one chunk. Therefore, memory limitation must be considered as one of the factors determining dialogue organization strategy.

Finally, this paper concentrates on the decision of turn taking unit to develop a dialogue organization strategy. However, it seems that the other levels of dialogue strategy such as exchange and move are also determined by the task context. We will investigate how the other levels of dialogue strategy are decided by these parameters.
Figure 6: A rough sketch of dialogue plan

**Conclusion**

This study has showed that the expert's linguistic behavior in an action instruction dialogue depends on the observable state of task execution and the discourse context. Based on the empirical results, we also give several plan selection preference rules and some criteria for determining the utterance content and the dialogue strategy for planning explanation dialogues.

**References**


