Comparing Error-Handling Strategies in Human-Human and Human-Robot Dialogues

Petra Gieselmann
Interactive Systems Lab
University of Karlsruhe
Am Fasanengarten 5
76131 Karlsruhe, Germany
petra@ira.uka.de

Abstract

In this paper error-handling strategies are evaluated as they can be found in human-human and human-robot dialogues. We compare human-human communication with unequal dialogue partners, such as foreigners, to human-robot communication in order to see how errors are indicated and finally repaired by the dialogue participants in both interaction types. The strategies used in human-human communication are ported to human-robot dialogues. The foreigner is taken as an example for a dialogue participant with less grammar and vocabulary coverage and can be seen as similar to the robot in this way. We finally compare the indicators for errors and the error repair strategies in human-human vs. human-robot dialogues.

1 Introduction

Today dialogue systems and humanoid robots talking to their human masters became very popular in the research community and some robots are already commercially available, such as Qrio from Sony or Asimo from Honda. Nevertheless, the human-robot communication is still not as natural as human-human communication so that it is one of the biggest challenges to develop a system which can cope with real world situations and can easily recover from errors.

Comparing the currently possible human-robot communication with the human-human communication we can see that in human-human communication we have efficient strategies to avoid errors and also to recover from them, such as for example *grounding* new information (Traum, 1999). Therefore, in this paper we want to evaluate and compare error handling strategies used in human-human vs. human-robot communication. We specifically focus on human-human dialogues with unequal dialogue partners where one dialogue partner has less grammar and vocabulary coverage than the other because this resembles to the situation we have in human-robot dialogues where the robot also has less grammar and vocabulary coverage than its human dialogue partner.

Our target scenario is a household situation, in which the user can ask the robot for help, such as setting the table, giving him some new recipes, opening the fridge and checking the food, switching certain lights on or off, getting some objects, such as cups, dishes, etc. (Gieselmann et al., 2003; Stiefelhagen et al., 2004). This context is specifically tailored for unexperienced and older users so that it is important that the user can talk to the robot in the same way as to a human servant. This means that the communication should be as natural and as comfortable as possible for the user and therefore, errors should be avoided or at least easy to correct, if they cannot be avoided beforehand.

This paper deals with error-handling strategies in task-oriented dialogues in human-robot as well as in human-human communication. Different aspects of misunderstandings and error handling are also explored, such as what it relates to, who notices and reacts to it and how. Section two gives an overview of related work in human-human communication. Communication theory and mechanisms for grounding and error recovery within interpersonal communication are explained. Section three deals with experimental details and results for human-robot communication within the household. Finally, section four gives a conclusion and an outlook on future work.

2 Human-Human Communication

2.1 Introduction

In this section, we want to analyse different aspects of misunderstandings and error handling, as they can be found in human-human conversation. First of all grounding is explained in detail. Then errors and repair strategies are explored by examining questions, such as who notices an error and who reacts to it in which way. Finally, we turn our attention to conversations with unequal dialogue partners and evaluate the indications for errors and the repair strategies.

2.2 Grounding

Grounding is an essential part of the communicative process establishing mutual knowledge between the participants in a dialogue: It concerns adding new information to the common ground of the dialogue participants (Traum, 1999; Poesio and Traum, 1998). Grounding can occur at the linguistic as well as at the cognitive level (Traum and Dillenbourg, 1998). Also non-verbal signals play an important role in ensuring an efficient communication.

The grounding criterion is reached according to Clark and Schaefer (Clark and Schaefer, 1989), when "the contributor and the partners mutually believe that the partners have understood what the contributor meant to a criterion sufficient for the current purpose". In human-human communication, we have efficient strategies for managing grounding issues, such as simple feedback strategies, paraphrasing, requesting clarifications or confirmations, etc.

In addition, Clark and Wilkes-Gibbs consider the *least collaborative effort* which tries to minimize the total effort of the collaborators (Clark and Wilkes-Gibbs, 1986) as another important principle. Therefore, in some cases the cost of producing a perfectly interpretable utterance might be higher than producing a flawed utterance, which can be easily repaired. This might result in problems in human-robot communication.

Nevertheless, a key behaviour of any artificial dialogue partner is its ability to engage in the grounding process with the user to ensure that everything said is mutually understood. For correcting misunderstandings, it is essential that both dialogue partners share the same knowledge and that a wrongly understood utterance can be replaced easily.

2.3 Errors and Repair Strategies in General

The analysis of errors in human-human dialogues is done by means of conversation analysis (Sacks et al., 1974; Atkinson and Heritage, 1984), where different kinds of dialogues are evaluated concerning the rules and procedures how an interaction takes place. According to the speech act theory of Searle (Searle, 1969), an *error* can be defined as not recognizing the intention of the speech act of the dialogue partner. In addition, we can find the following indicators for errors (Bremer, 1997; Marti, 2001):

- Indicating non-understanding
- Questioning for additional information
- Repeating a central element
- Implicit indicators: Ignoring the error or canceling

Furthermore, we can distinguish two kinds of errors in dialogues: *Non-understanding* vs. *misunderstanding*. Non-understanding means that one of the dialogue partners cannot find any information in the utterance of the other. In human-robot dialogues, such non-understandings can be due to the fact that the grammar does not cover the user utterance which cannot be parsed therefore. Also on the pragmatic level, non-understanding is possible, when the user utterance is inconsistent with the current discourse within the dialogue manager and can therefore not be integrated.

Misunderstanding is much more common in human-human dialogues so that one dialogue partner misinterprets the utterance of the other dialogue partner. In human-robot dialogues this means that a user utterance can be parsed and the semantic interpretation is integrated in discourse, but does not correspond to the user's intention. This is above all due to speech recognition errors, ie. a word has been misrecognized. But also a semantic misunderstanding might be possible, if some information from the user utterance has been integrated wrongly in the existing discourse.

To repair these two kinds of errors, we can find different strategies in human-human communication, as explained by Schegloff and Jefferson (Schegloff and Jefferson, 1990): On one hand, the repair can be self-initiated vs. other-initiated and on the other hand, the execution of the repair can be done by the speaker himself or by the dialogue partner (self-correction vs. other-correction), as you can see in table 1.

Nevertheless, the dialogue partners decide on their own whether an utterance or some part of an utterance are a trouble source and need to be repaired. Not every error has to be repaired, if the communication is not hampered by this error. Either not every reparation is seen as an error by the dialogue participants.

	self-repair	other-repair
self-initiated	self-rectifications	request for help
other-initiated	clarification questions	other-initiated other-repair

Table 1: Error Repair Strategies in Human-Human Conversation

2.4 Errors in Dialogues with different Dialogue Partners

Errors in human-human dialogues and the resolution of these errors are of special interest for dialogues with different, unequal communication partners such as children and foreigners because in the human-robot dialogue we also have inhomogeneous partners. All these communication partners have problems within the communication, but nevertheless they are able to pursue their communication goal by means of metalinguistic and metacognitive abilities (Perlis and Purang, 2001).

As the typical case of error repair in dialogues with inhomogeneous partners we can find other-initiated self-repair (See Table 1) because the speakers need the feedback from the dialogue partner whether they have been understood or not (Marti, 2001). This means that the less competent participant indicates that he did not understand and the other participant tries then to find another formulation. Transferring the situation to our robot scenario, this means that the robot has to indicate to the human participant whenever it did not understand and the human would then reformulate his utterance.

In a user study (See section 3), we evaluate whether this statement of mostly other-initiated self-repair is also true for communication with an artificial partner, such as a robot. In addition, we have a look at the indications of errors and the strategies used to repair errors.

The non-understanding can concern single words or also complex situations. In human-human dialogues with foreigners, single words are just indicated as being unknown by repeating them (Bremer, 1997). The dialogue partner then tries to correct or repair this error by explaining the word or reformulating the whole expression. Therefore, the challenge for the dialogue manager is at first to detect an error and then to initiate a repair dialogue.

2.5 Error Recovery

In human-human dialogues, whenever a dialogue tends to fail, humans use the following strategies to prevent this (Faerch and Kasper, 1983; Tarone, 1981; Marti, 2001):

- Achievement strategies
  - Paraphrases
  - Restructuring the whole sentence
  - Approaching resp. isolating the missing information
  - Repetition
  - Metalinguistic transfer
  - Ask for help
- Functional reduction strategies
  - Functional reduction of the intention
  - Giving preselected answers to choose from
  - Changing the theme
- Formal reduction strategies
  - Morphosyntactic reduction
- Ratification
  - Approval
  - Repetition of the indicator

In detail, this means that whenever the users apply achievement strategies, they explain the semantics of the utterance by extended communicative means. For example, they reformulate their utterances or switch to another language. Functional reduction strategies change the original meaning. The users reduce their communicative goal to prevent any more problems. They replace their original intention by another one. Formal reduction strategies represent a simplification of the grammar and/or vocabulary. Ratifications serve as confirmations to the preceding utterance.

These strategies can also be used in combination. For example, paraphrases and restructuring of the whole sentence are often applied at the same time. Also restructuring and morphosyntactic reduction are sometimes combined within the same utterance.

Noteworthy, these strategies are used independently from the grammatical competence in a language. This means that it is important that our dialogue system is also able to use such strategies in a similar way as humans do and recognize them when such strategies are used by the human dialogue partner.

3 Experimental Details

3.1 User Study

We made a small user study with four users and evaluated the human-robot dialogues in a black box evaluation. This means that we analysed the utterances of the users and the robot and did not evaluate whether the user utterance has been recognized correctly by the speech recognizer or whether it could be parsed correctly or transformed to the correct semantics. In this way, we want to assure an objective evaluation from an outside point of view and not from a developer's perspective.

We told the users that they got a new household robot who can help them in the kitchen by setting the table, giving them some recipes, switching certain lights on or off, getting some objects, looking in the fridge, etc. and that they can now play with it. We did not give them any detailed instructions on how to speak to the robot, but let them freely explore. This results in more errors to be analysed because the participants sometimes mentioned tasks the robot cannot do at all or formulations not covered by the grammar at the moment. In addition, in this way we can assure that we cover the user needs in our black box evaluation.

3.2 Error Repair Strategies

All together, we had 212 turns, 106 by the user and 106 by the robot. Out of the 106 robot turns, there were 26 turns with no reaction from the robot. We counted these cases also as errors of the robot because it did not give any answer to the user.

There were 71 errors. As expected, the results showed that there were most of the time other-initiated self-repairs (94.36% of all errors were corrected in this way). This means that the robot indicated in one or another way to the user that it did not understand anything and the user tried to reformulate his utterance.

In the other cases (5.74%), the user changed the topic of the conversation unexpectedly and did not answer the preceding question from the robot. This was in all cases due to an also unexpected question from the robot's side which the user ignored in this way.

Indicators for Errors	Rate
Indicating non-understanding	24.0%
Questioning for add. information	8.0%
Repeating a central element	5.33%
Ignoring the error or canceling	62.67%

Table 2: Indicators for Errors in Human-Robot Conversation

3.3 Indicators for Errors

We evaluated the indicators for errors by the robot as well as by the user to assure that the robot recognizes whenever a repair dialogue starts and that it uses similar mechanisms to start a repair dialogue. Since we had other-initiated self-repair most of the time, the indicators for non-understandings can be mostly found in the robot's utterances.

Indicators for errors can be found on one hand in the prosody of the user utterance, since users often hyperarticulate corrections, as already explained by different researchers (Swerts et al., 2000; Soltau and Waibel, 2000; Oviatt, 1998). On the other hand, we can also find indicators for errors in the recognized words. As you can see in table 2, in 62.67% of all errors, we found utterances ignoring the error and starting another topic. Half of these utterances were utterances where the robot did not react at all to the preceding user utterance, but completely ignored it and did not utter anything. In about 24% of the errors, the robot indicated the non-understanding by unspecific questions, such as "How can I help you?". Sometimes (8.0%), the robot asked for additional information by questions, such as "Where should I go?", "Which cup do you want me to take?". In only 5.33%, the robot just repeated its previous question.

In human-human communication with foreigners, we find a more balanced distribution of indicators for errors with approximately 22% in all the four categories (Marti, 2001). Therefore, in the future we will adapt the robot to a more explicit error indication to the user, as it is used in human-human communication.

In addition, the user reactions to the different indicators for errors showed that the users preferred a more explicit error indication: When the robot just ignored the preceding user utterance, the users felt lost and tried different strategies, such as paraphrasing, switching to another task, asking the robot what it can do, etc. This kind of error indication did therefore not result in an immediate error recovery in any single case in the user test. Also indication of non-understanding with unspecific questions resulted in user confusion because the users did not know what the robot did not understand in detail and only in 16% of these cases, the errors could be resolved after the indication and the robot finally understood what the user wanted it to do.

On the other hand, errors can be resolved successfully in all the cases, where the robot asked a specific question to the user for additional information, and the user answered this question. The only problem with the specific questions was that if the users did not expect such a question at this time of the conversation, they did not answer it and therefore the error could not be resolved. In all the cases, where the robot repeated its preceding question, the users paraphrased their last answer. In half of the cases, the communication could be put back on track successfully in this way. In all the other cases, the users again used a formulation which has not been covered by the current grammar and has therefore not been understood by the robot either. To conclude, these results strengthen the importance of clarification questions from the robot to support the user during the error recovery dialogue.

3.4 Strategies for Error Recovery

Most of the error recovery is done by the human at the moment, given the fact that we mostly had other-initiated self-repair. Not all the errors are corrected, some are just ignored. We found in 37.3% of all user utterances achievement strategies for error recovery, as you can see in table 3. Paraphrases and reformulations are very common (14.67% of the user utterances). Restructuring and approaching the missing information can be found in only 9.33%. Sometimes, the users even repeated an utterance exactly in the same way as in the preceding utterance (5.33% of all user utterances) hoping that the robot would understand them now.

Furthermore, we found in 8% of all the user utterances correcting utterances which were not covered in the original models used in human-human communication (See section 2.5). Such utterances were used, when the robot said something which did not cover the user intention at all or even represents the opposite of the user intention.

Functional reduction strategies consist most of the time of changing the theme completely. We found only one case with a functional reduction of the intention, as you can see in the lines 11 - 13 in the example dialogue (See Figure 1), where the user asked the robot to bring him a cup and since the robot did not understand him, changed his intention and made the robot first go to the place where the cup is.

Error Recovery Strategy	Rate
Achievement strategies	37.3%
Functional reduction strategies	16.0%
Formal reduction strategies	4.0%
Ratification	2.7%

Table 3: Strategies for Error Recovery in Human-Robot Conversation

There were only very few examples of formal reduction strategies. Mostly, the users even reformulated their utterances in a morphosyntactically more complicated way, as you can see in the example dialogue in the lines 21 - 23 (See Figure 1).

We found only very few examples of ratification which might be due to the fact that the whole dialogue is task-oriented and the users see when the robot finally executes the task they asked it for. Therefore, no more explicit approval is needed.

Also in human-human dialogues achievement strategies are used mostly (Marti, 2001). However, human-human communication seems to be more efficient in error repair which might be due to the fact that the paraphrases and reformulations are still not understood by the system and that repetitions by the user are hyperarticulated so that the recognition is even harder (Soltau and Waibel, 2000), as might be explored in future studies.

3.5 Example Dialogue

The example dialogue in figure 1 starts with some achievement strategies applied by the user (lines 1-6). He wanted to get some orange juice, but the robot just informed him that there is juice in the fridge. Therefore, he tried to isolate the missing information and explicitly asked, whether the available juice is orange juice or apple juice (See line 5). Since the robot seemed not to understand this question either, the user changed to another topic and asked the robot then to bring him a cup (Changing the theme).

Since the robot only understood some part of the user intention, the user explained which cup he wanted to get by repeating his utterance and also

1. User: Robbi, do we have any orange juice?
2. Robbi: I will have a look at the fridge. There are tomatoes, apples, noodles, eggs, milk, butter, cucumbers, coke and juice.
3. User: What about orange juice. Do we have any orange juice?
4. Robbi: With this stuff in your fridge you can make different dishes, such as spaghetti napoli, spaghetti pesto, a pancake and a salad. For which recipe are you looking?
5. User: Robbi, is it orange juice or apple juice?
6. Robbi: I will have a look at the fridge. There are tomatoes, apples, noodles, eggs, milk, butter, cucumbers, coke and juice.
7. User: Please bring me this cup (+ Pointing Gesture)
8. Robbi: Which cup do you want me to take?
9. User: This cup. (+ Pointing Gesture)
10. Robbi: Where should I put it?
11. User: Please bring me this cup.
12. Robbi: –
13. User: Go.
14. Robbi: How much should I go forward?
15. User: Now bring me the blue cup.
16. Robbi: Where should I go?
17. User: Bring me the blue cup.
18. Robbi: Where should I go?
19. User: Bring me the yellow cup.
20. Robbi: Going to bring you the yellow cup.
21. User: Bring me some food.
22. Robbi: –
23. User: I asked you to bring me some food.
24. Robbi: How can i help you?
25. User: Some food please.
26. Robbi: Where do you want me to put it?
27. User: I would like to eat a salad.
28. Robbi: For a salad you need tomatoes, cucumbers, and dressing.
29. User: Thanks.
30. Robbi: You are welcome.

Figure 1: Example Dialogue

using deictic pointing gestures. Since the robot still did not understand, the user finally changed his intention and wanted to make the robot first go to the place where the cup is (Formal reduction).

In lines 17 - 20, we have a nice example of a functional reduction where the user first wants the robot to get the blue cup and then changed to the yellow cup, as he noticed that the robot did not understand him after he repeated his utterance.

As already explained in the previous section, there were very few morphosyntactic reductions, but mostly the utterances got even more complicated when paraphrased. In the example dialogue in the lines 21 - 28, the robot did not understand the user intention of getting some food for several turns. The user first tried a more complex paraphrase ("I asked you to bring me some food.") and after that a morphosyntactic reduction ("Some food, please."). Finally, he switched to a more concrete intention saying "I would like to eat a salad."

An example for a ratification can be found in the last two lines: The robot told the user the ingredients of a salad and the user thanked it to confirm that it understood his intention correctly.

Generally speaking, as you can also see in this example dialogue, lots of errors are just not repaired, but the user made the robot accomplish another task (See for example line 7). This might also be due to the fact that the users did not have predefined tasks to accomplish, but could decide on their own what they want to make the robot do.

4 Conclusion & Outlook

In this paper, we presented error handling strategies used in human-human dialogues and evaluated whether these strategies are also used in human-robot communication. We focused on human-human communication with unequal partners with different vocabulary and grammar coverage, such as foreigners, because the results can be adapted to human-robot dialogues easily. We compared the indicators for errors and the error repair strategies in a human-human vs. human-robot communication.

At the moment, the most common indicator for an error in human-robot communication is a sudden change of the current dialogue topic. This often results in problems for the human user who has no idea what exactly has not been understood. Therefore, in the future we want to improve the error indications from the robot and adapt them to the ones used in human-human communication so that more explicit error indications, questioning for additional information and repeating a central element are used. This is of special importance given the fact that the robot is the less competent dialogue partner and therefore relies on its human dialogue partner to correct any error as soon as possible.

The results revealed that the participants mostly used achievement strategies to recover from errors. Also the functional reduction strategy was used several times which might be due to the fact that the users did not have predefined tasks to solve and could therefore easily abandon a task and start a new one.

In the future, we want to use the results of this user study to enhance the error handling strategies of the robot. We will use confidence measures from the speech recognizer for example so that the robot can explicitly indicate what it did not understand. In this way, the robot can also repeat specific words with very low confidences to indicate an error and start an error recovery dialogue.

5 Acknowledgements

This work was supported in part by the German Research Foundation (DFG) as part of the SFB 588 on humanoid robots and by the European Commission under project CHIL (contract #506909).

References

J. M. Atkinson and J. Heritage. 1984. *Structures of Social Action. Studies in Conversation Analysis*. Cambridge UK: Cambridge University Press.

K. Bremer. 1997. *Verstaendigungsarbeit: Problembearbeitung und Gespraechsverlauf zwischen Sprechern verschiedener Muttersprachen*. Tuebingen: Niemeyer.

H. H. Clark and E. F. Schaefer. 1989. Contributing to discourse. *Cognitive Science*, 13:259–294.

H. H. Clark and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.

C. Faerch and G. Kasper. 1983. Plans and strategies in foreign language communication. *Faerch, C. and Kasper, G. (eds.) Strategies in Interlanguage Communication*, pages 20–60.

P. Gieselmann, C. Fügen, H. Holzapfel, T. Schaaf, and A. Waibel. 2003. Towards multimodal communication with a household robot. *Proceedings of the Third IEEE International Conference on Humanoid Robots (Humanoids)*.

J. Marti. 2001. *Verstehensschwierigkeiten unter Nicht-MuttersprachlerInnen. Eine empiriche*

Studie anhand von Asylsuchenden in Bern. Ph.D. thesis, University of Bern.

S. Oviatt. 1998. The cham model of hyperarticulate adaptation during human-computer error resolution. *Proceedings of the International Conference on Spoken Language Processing (ICSLP'98)*.

D. Perlis and K. Purang. 2001. Conversational adequacy: Mistakes are the essence. Technical report, Department of Computer Science, University of Maryland.

M. Poesio and D. Traum. 1998. Towards an axiomatization of dialogue acts. *Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues (13th Twente Workshop on Language Technology)*, pages 207–222.

H. Sacks, E. Schegloff, and G. Jefferson. 1974. A simple system for the organization of turn-taking im conversation. *Language*, 50(4):696–735.

E. Schegloff and G. Jefferson. 1990. The preference for self correction in the organization of repair in conversation. *Psathas, G. (Ed.): Interaction competence*.

J.R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, UK.

H. Soltau and A. Waibel. 2000. Specialized acoustic models for hyperarticulated speech. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP'00)*.

R. Stiefelhagen, C. Fuegen, P. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel. 2004. Natural human-robot interaction using speech, gaze and gestures. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

M. Swerts, J. Hirschberg, and D. Litman. 2000. Corrections in spoken dialogue systems. *Proceedings of the International Conference on Spoken Language Processing (ICSLP'00)*.

Elaine Tarone. 1981. Some thoughts on the notion of communication strategy. *TESOL Quarterly*, 15(3):285–295.

D. R. Traum and P. Dillenbourg. 1998. Towards a normative model of grounding in collaboration. *Working notes of the ESSLLI-98 workshop on Mutual Knowledge, Common Ground and Public Information*.

D. R. Traum. 1999. Computational models of grounding in collaborative systems. *Psychological Models of Communication in Collaborative Systems - Papers from the AAAI Fall Symposium*, pages 124–131.