

# Relating Initial Turns of Human-robot Dialogues to Discourse

## **Maxim Makatchev**

Robotics Institute  
Carnegie Mellon University  
Pittsburgh, PA, USA  
[mmakatch@cs.cmu.edu](mailto:mmakatch@cs.cmu.edu)

## **Min Kyung Lee**

Human-Computer Interaction Institute  
Carnegie Mellon University  
Pittsburgh, PA, USA  
[mklee@cs.cmu.edu](mailto:mklee@cs.cmu.edu)

## **Reid Simmons**

Robotics Institute  
Carnegie Mellon University  
Pittsburgh, PA, USA  
[reids@cs.cmu.edu](mailto:reids@cs.cmu.edu)

## ABSTRACT

User models can be useful for improving dialogue management. In this paper we analyze human-robot dialogues that occur during uncontrolled interactions and estimate relations between the initial dialogue turns and patterns of discourse that are indicative of such user traits as persistence and politeness. The significant effects shown in this preliminary study suggest that initial dialogue turns may be useful in modeling a user's interaction style.

**Categories and Subject Descriptors:** H.5.2 [Information Interfaces and Presentation]: User Interfaces | natural language

**General Terms:** Experimentation, Human Factors

**Keywords:** Human-robot interaction, dialogue

## 1. INTRODUCTION

In many human-robot interaction scenarios, the only information available to the robot about the particular user is what can be obtained from the single instance of the ongoing interactive session. For example the Roboceptionist [1, 2], installed at a high-traffic entrance of a university building, does

not track users from session to session. Therefore, the user model has to be constructed on-the-fly, ideally from the first turns of the dialogue, so that that dialogue manager can take advantage of the model in adapting to the user while the interactive session is still in progress.

In this study, we analyze transcripts of human-robot dialogues with the goal of predicting dialogue patterns that potentially indicate user's interaction styles from the first turns of the dialogues. In particular, we are interested in predicting such user traits as the ability to carry on after robot demonstrates lack of understanding (persistence), as well as the user's adherence to social norms, such as thanking the robot after the robot gave an answer to the user's question (politeness) and ending an interaction with a farewell. We present preliminary results of predicting these traits based on whether the interaction has been initiated by the robot and whether the user's first turn was a greeting. Our analysis shows, for example, that users starting with a greeting are more likely to have their questions answered, exhibit more persistence, and are more than 3 times as likely to be polite. The paper concludes with a discussion of results and an outline of the ways they can be used in the dialogue management.

## 2. HUMAN-ROBOT DIALOGUE CORPORA

The corpus of human-robot dialogues represents transcripts of uncontrolled interactions that are collected on a near daily basis. Users type their input on a keyboard mounted in front of the robot and the robot's turns are spoken. The transcripts are automatically segmented into individual dialogues, and dialogues with more than 20 turns are discarded to eliminate some of the outliers of the segmentation procedure. We have manually labeled 1382 turns of 196 dialogues that occurred over 6 days in March of 2008 with respect to such dialogue acts as greeting, thanks, bye (farewell), question, answer, dismissal (robot's admitting its inability to understand the preceding user's turn), and rude language. Discourse patterns of our interest, such as question answered, persistence, and politeness are expressed in terms of these dialogue acts.

We use this manually labeled corpus of dialogues to train the decision tree classifiers for each of the dialogue acts. Using unstemmed words as the features, these classifiers achieve accuracy of at least 89% (10-fold cross-validation is used to select the size of the trees). The high accuracy of the automated labeling justifies expanding the analysis to the larger corpus of dialogues. The results presented below correspond to the automatically labeled corpus of 1676 dialogues (11024 dialogue turns) that occurred over the months of March and April of 2008.

## 3. EXPERIMENTS

In the following experiments, we estimate the relation between the two (not mutually exclusive) ways to begin a dialogue and the discourse: (1) whether the dialogue has been initiated by the robot and (2) whether the user has started the dialogue with a greeting (e.g. \Hi", \Good morning"). We define a dialogue as initiated by the robot if the user started typing within 10 seconds from the time when robot has greeted a passer-by. Greeting of a passer-by is triggered by a user entering an area that is close to the robot (as sensed by a laser range scanner) with a minimal forward velocity.

The features of dialogues that we compare include start time, dialogue duration in seconds, dialogue duration in number of turns, total number of words, average number of words per turn, presence of user's farewell, robot's admitting its inability to understand the preceding user's turn (dismissal), user's rude language, user's persistence (robot's dismissal followed by user's turn that is not farewell), user's question, user's question answered (not dismissed), and politeness (user's thanking the robot after the question has been answered). The features under user greeting/no-greeting conditions also include total number of words and average number of words per turn for the "inner" dialogue turns that exclude an initial greeting and trailing farewell turns. This metric of verbosity can be useful in that it excludes the two turns trivially affected by the presence of the greeting: the greeting turn itself and the farewell turn, which is highly correlated with the greeting (Table 2).

The results are shown in Tables 1 and 2. Where the units are not specified, the numbers represent fractions of all applicable dialogues. Results in bold denote the differences significant at  $\alpha = 0:05$  according to two-sample t-test.

*Table 1: Relation between the initiator of the dialogue and the discourse, using dialogue turns labeled by a classifier.*

classifier	robot-init.	human-init.	p-value
start time	2:29pm	2:43pm	0.1563
duration (sec)	38.95	47.77	0.379
num. of turns	6.34	6.70	0.148
num. of words	9.10	9.28	0.6966
words per turn	2.70	2.70	0.986
greeting	<b>0.42</b>	<b>0.37</b>	<b>0.0338</b>
bye	<b>0.12</b>	<b>0.18</b>	<b>0.0011</b>
robot's dismissal	0.51	0.53	0.3758
rude	0.02	0.02	0.7176
persistent	0.72	0.70	0.6037
user's question	0.58	0.61	0.2222
question answered	<b>0.52</b>	<b>0.42</b>	<b>0.0029</b>
polite	0.14	0.17	0.5021

*Table 2: Relation between a greeting in the user's first turn and the discourse, using dialogue turns labeled by a classifier.*

classifier	robot-init.	human-init.	p-value
start time	2:24pm	2:47pm	0.0238
duration (sec)	43.61	45.44	0.8438
num. of turns	7.69	5.88	< 0:0001
num. of words	10.25	8.56	0.0002
words per turn	2.35	2.92	< 0:0001
num. of words (inner)	8.47	8.30	0.6949
words per turn (inner)	3.20	2.98	0.0122
bye	0.20	0.13	< 0:0001
robot's dismissal	0.46	0.56	< 0:0001
rude	0.02	0.02	0.5558
persistent	0.77	0.67	0.0014

classifier	robot-init.	human-init.	p-value
user's question	0.62	0.59	0.3420
question answered	0.50	0.43	0.0232
polite	0.25	0.08	< 0:0001

Earlier work has shown a significant effect of time of day on the duration of interactions with a previous version of the Roboceptionist (Valerie) [2]. Although our results show that the interactions where the first user's turn is a greeting tend to occur on average 23 minutes earlier than the interactions where user starts with a dialogue act other than a greeting, no other significant effects on the discourse features were discovered. We intend to incorporate the effect of time of day in our analysis in a future work.

While the initiator of the dialogue does not show as much effect on the discourse as whether the user started with a greeting, robot-initiated dialogues show a slight increase in being understood by the robot is not explained by the presence of trivial "hi-hi-bye-bye" interactions, since the difference remains significant for interactions containing at least 4 and at least 6 turns.

## 4. CONCLUSIONS AND FUTURE WORK

The presence of significant effects of user's greeting on such aspects of dialogues as persistence and politeness can be helpful in user modeling and can potentially be exploited by a dialogue manager. For example, if there is an indication that the user is less persistent, the robot could take extra care in verbalizing dismissals, with the goal of increasing the chance that the user will rephrase their previous turn. Similarly, a lack of thanks after the robot's answer from the user that is expected to be polite may serve as feedback that the answer to the user's question was not relevant or helpful. The robot might also have a longer delay before considering an interaction to be over due to inactivity, if the user has not yet provided an expected farewell.

Our study demonstrates the potential of using initial dialogue turns to predict discourse features that indicate a user's interaction style. A possible further work is to use more sophisticated features of the dialogue turns, for example content words [3], and to incorporate the time of day.

## 5. ACKNOWLEDGMENTS

The authors would like to thank Brett Browning for useful discussions and Antonio Roque for his helpful comments. This publication was made possible by the support of an NPRP grant from the Qatar National Research Fund. The statements made herein are solely the responsibility of the authors.

## 6. REFERENCES

- [1] R. Gockley, A. Bruce, J. Forlizzi, M. Michalowski, A. Mundell, S. Rosenthal, B. Sellner, R. Simmons, K. Snipes, A. C. Schultz, and J. Wang. Designing robots for long-term social interaction. In *Proc. Int. Conf. on Intelligent Robots and Systems*, pages 2199-2204, August 2005.
- [2] R. Gockley, J. Forlizzi, and R. Simmons. Interactions with a moody robot. In *Proc. Int. Conf. on Human-Robot Interaction*, pages 186-193, March 2006.

[3] A. Purandare and D. Litman. Content-learning correlations in spoken tutoring dialogs at word, turn and discourse levels. In *Proc. Int. FLAIRS Conf.*, May 2008.