

Dialogue Patterns of an Arabic Robot Receptionist

Maxim Makatchev[‡], Imran Fanaswala[†], Ameer Abdulsalam[†], Brett Browning^{†‡},

Wael Ghazzawi[†], Majd Sakr^{†‡} and Reid Simmons[‡]

[†]Carnegie Mellon University in Qatar, Doha, Qatar

Email: {imranf,ameer,wghazzaw,msakr}@qatar.cmu.edu

[‡]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

Email: {mmakatch,brettb,reids}@cs.cmu.edu

Abstract—Hala is a bilingual (Arabic and English) culturally-sensitive robot receptionist located at Carnegie Mellon University in Qatar. We report results from Hala’s deployment by comparing her English dialogue corpus to that of a similar monolingual robot (named “Tank”) located at CMU’s Pittsburgh campus. Specifically, we compare the average number of turns per interaction, duration of interactions, frequency of interactions with personal questions, rate of non-understandings, and rate of thanks after the robot’s answer. We provide possible explanations for observed similarities and differences and highlight potential cultural implications on the interactions.

Keywords: Human-robot interaction, dialogue, culture, Arabic language, English language

I. INTRODUCTION

Studies across multiple cultures suggest that individuals and groups are more inclined to like those characters that they perceive as culturally similar to themselves [1]. We designed Hala as a bilingual (English and Arabic) version of a robot receptionist, conspicuously placed in the reception area of CMU-Qatar. The goal of the project is to explore culturally-aware human-robot interaction.

The robot’s face is rendered on a flat-screen that is mounted on a neck joint. Users interact with Hala, using a keyboard placed in front of her, in one of three available input modes: English, Arabic or transliterated Arabic (3arabi). Depending on the input mode, Hala will respond in English or Arabic, by producing a synthesized voice reply as well as text that appears next to her face.

Hala is designed to provide information about campus directions, weather, local events and answer queries regarding her personal life. Her backstory is that of a young unmarried Arab female. The robot uses a rule-based dialogue manager that consists of a knowledge base of canned utterances and rules that trigger responses in English and Arabic. At present, the English knowledge base is significantly broader in terms of its content coverage.

Tank is a similarly designed robot receptionist located at a high-traffic entrance of a CMU building since 2004 [2].

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.



Fig. 1. Hala is situated in a large open area amongst human receptionists.

Keeping aside hardware similarities, their personalities, environment and user population are different. Tank only speaks English, and his personality is of a retired CIA agent. Hala is situated at the reception counter next to human receptionists and security personnel (Figure 1). Tank is located inside a kiosk and does not have a human receptionist in his vicinity. CMU-Qatar hosts an international community, but demographics lean towards an Arab, South and South-East Asia populace. On the other hand, CMU-Pittsburgh has a typical mix of communities found in a North-American university.

Given the interactions recorded by Tank and Hala, we want to determine how these interactions are influenced by the above mentioned factors. In this study we compare the robots’ dialogue corpora, with respect to basic interaction statistics such as duration and average number of turns per interaction, and with respect to such discourse features as a user greeting the robot, user asking the robot a question, and user thanking the robot after it attempted to answer user’s question.

II. HALA’S INTERACTIONS

An interaction is a sequence of consecutive user and robot actions (turns). While turns consist of utterances, GUI actions of switching between input modes and also the Roboceptionist’s non-verbal actions (head nods and facial expressions), in this study we only focus on utterances and thus use the terms “dialogue” and “interaction” interchangeably. The

robots employ strict turn-taking, hence a user and the robot have equal number of turns per interaction. Hala’s interaction boundaries are defined by timing out, while Tank uses a combination of a timeout and a laser range scanner to detect when a person has left his vicinity.

We analyzed six weeks of Hala’s interactions that took place in October and November of 2009, 2 months after the robot’s initial deployment. Over this period she conducted on average 18.3 interactions per day, totalling 803 interactions. Tank had 36.4 interactions per day over the period of 6 weeks in March and April of 2008, with a total of 1676 interactions. The majority of Hala’s interactions were conducted in English at 89%, with only about 6.1% of interactions containing Arabic input (in either of Arabic input modes). The rest of the interactions do not contain utterances—users just played with the input mode buttons. The scarcity of dialogues that contained Arabic utterances could be explained by the robot’s limited Arabic coverage at present (and hence, a high rate of non-understandings). Notably, a number of Hala’s dialogues typed in English mode contained Arabic greetings and other ritualistic expressions. All of Tank’s interactions were conducted in English.

A. Hala’s and Tank’s Interactions in English

Due to differences with respect to their knowledge base and dialogue rule coverage, a direct comparison between their corpora is difficult. Nevertheless we attempt to conduct such a comparison, and normalize Hala’s corpus by focusing on the subset of the dialogues that were conducted entirely in English input mode. We annotated both corpora’s dialogue turns with labels corresponding to user’s greeting *U:Greeting*, user’s question *U:Question*, robot’s answer (excluding non-understanding) *R:Answer* and user’s thanks *U:Thanks*, using dialogue act the classifiers that were trained on Tank’s data and manually adapted to the observed peculiarities of Hala’s dialogues (e.g. using Arabic “salaam” in dialogues typed in English mode). The turn labels we use are a small subset of hierarchical dialogue act ontology DIT++ [3] adapted to suit our domain. Each dialogue turn may have multiple labels. A sequence of labels corresponding to adjacent dialogue turns can form a pattern, such as *AnswerThanked* (*R:Answer* followed by *U:Thanks*). We also assign labels *SomeAnswerThanked*, *SomeUserQuestion* to entire interactions that contain at least one of the corresponding patterns, and label *FirstTurnGreeting* to interactions that begin with a user’s greeting. In Table I we compare means and standard errors (SE) of (a) interaction durations in seconds and in number of utterances, (b) fractions of interactions that start with user’s greeting, contain user’s question, or user’s thanks after the robot’s answer, and (c) fractions of questions answered by the robot and fractions of those answers that were thanked by the user.

B. Discussion

The comparison shows that Hala’s dialogues last almost twice as long as Tank’s dialogues and on average contain an extra pair of utterances. The differences in the fraction

TABLE I
TRENDS OF HALA’S AND TANK’S INTERACTIONS IN ENGLISH.

	Hala		Tank	
	mean	SE	mean	SE
duration, sec	120	6	63	6
number of utterances	11.10	0.48	8.9	0.24
<i>FirstTurnGreeting</i> , %	38.7	1.8	39.4	1.2
<i>SomeUserQuestion</i> , %	69.5	1.7	60.3	1.2
<i>QuestionAnswered</i> , %	79.8	0.9	26.8	1.0
<i>AnswerThanked</i> , %	2.3	0.3	12.9	1.4
<i>SomeAnswerThanked</i> , %	6.6	1.2	15.8	1.8

of interactions that begin with a user’s greeting or contain a question are small. The fraction of Tank’s answers that are thanked is almost 6 times greater than that of Hala. Consequently, a larger fraction of Tank’s interactions (15.8% vs Hala’s 6.6%) that contain the robot’s answer also contain at least one instance of user’s thanks.

One likely explanation for the lower rate of user thanks lies in the nature of the questions. About 57% of Hala’s (and only 26% of Tank’s) interactions contain personal questions. These questions range from inquiries about her marital status to questions about the objects in her vicinity on the receptionist counter. Answering such personal questions may not constitute a thankable action [4], since in human-human dialogues such personal information exchange often serves the role of a social act that benefits the whole interaction rather than one of the interlocutors.

III. CONCLUSION

In spite of similar hardware, the two robots in our study expose significant variation in their dialogue patterns. These arise due to differences in robot personalities, coverage of their knowledge bases, their surroundings and cultural variation in their community of users—or some combination thereof. Users tend to thank Hala less frequently, likely due the observed bias towards personal questions. This bias can be alleviated or exploited, for example by allowing the robot to take initiative in choosing a conversation topic or by increasing the coverage within the popular topics, respectively.

We also observed frequent cases of code-switching, when users typed Arabic expressions like “salaam alaikum” (“peace be upon you”) within an English dialogue. This suggests that the robot’s English and Arabic language processing facilities should be simultaneously active for any text input mode.

REFERENCES

- [1] L. M. Osbeck, F. M. Moghaddam, and S. Perreault, “Similarity and attraction among majority and minority groups in a multicultural context,” *Int. J. of Intercultural Relations*, vol. Vol. 21, no. 1, pp. 201–210, 1997.
- [2] 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*, August 2005, pp. 2199–2204.
- [3] H. Bunt, “Dimensions in dialogue act annotation,” in *Proc. Int. Conf. on Language Resources and Evaluation (LREC)*, 2006.
- [4] K. P. Schneider, “Genre matters: Textual and contextual constraints on contemporary English speech behaviour,” *Anglia — Zeitschrift für englische Philologie*, vol. 125, no. 1, pp. 59–83, 2007.