NLG for Situated Human-Robot Dialogue
Statement of Interest for the NLG for HRI Workshop, HRI 2020

Ivana Kruijff-Korbayová, DFKI, Saarbrücken, Germany, ivana.kruijff@dfki.de

1 PREFACE

I am interested in the NLG for HRI Workshop at HRI 2020 for a variety of reasons. To begin with, I have been interested and on-and-off working on NLG for many years, starting with the generation of instruction texts [1], then NLG for dialogue systems [2, 3] and finally NLG and dialogue in the context of social HRI [4]. I have worked on the more subtle issues of verbal output realization, such as using information structure to control word order [5] and/or intonation [6], in order to have utterances fit well into the context. I/We have used the "classical" approach of rule-based utterance planning and grammar-based generation. In our work on speech-based HRI for sustaining a long-term relationship, we have focused on generating a large variety of verbal output realizations and on conveying familiarity [7], using a combination of grammar-based and template-based generation. In the work I have done so far, context has almost always been linguistic. That is where I feel I have not yet quite made the step into the true area at the intersection of language/speech and robotics. This is the direction I want to go in the near future: *situated NLG*, by which I mean generating language which refers to the physical situation context. As a side note, I like to distinguish situated language/interaction from multimodal interaction (which may or may not refer to the physical situation) and from embodied interaction (where the issue is which modalities, i.e., kinds of signals, of the "body" are being used to interact). Our human-robot dialogue system uses the VonDA architecture, cf. Fig. 1 [9].

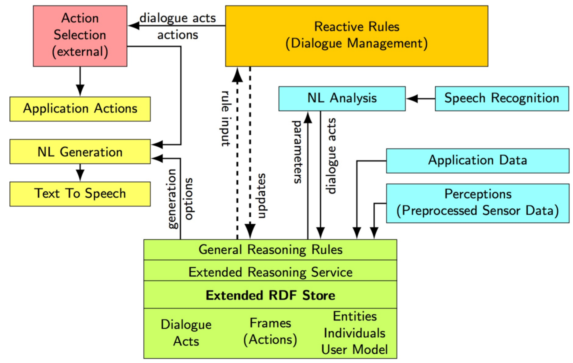

Figure 1: Human-robot dialogue system architecture

2 CURRENT PROJECTS

Social HRI in public spaces

In the project INTUITIV[1] we are studying intuitive-nonverbal and informative-verbal robot-human communication on use-cases set in a rehabilitation clinic environment: giving *indoor route directions* while accompanying clients to places in the cilnic; handling *incidental encounters* with bystanders and bypassers; and *passing objects*, e.g., passing bandages to a nurse. The robots we use include an autonomous walking aid, a ground platform that is to be used to transport luggage and a robotic arm, cf. Fig.2. The overall aim is that the robots exhibit behaviours that help to establish understanding and trust for patients and caregivers in the actions of the robotic system. We study the need for linguistic communication and its appropriate realization in various situations, together with the influence of iconic information in the form of sounds, displays, simulated eye movements, etc. on interpretability. Our goal is a situation-adapted selection an appropriate interaction possibility and combination of means of realization.

For the first prototype of indoor route directions generation we have started from semantically annotated maps and template-based NLG. We are first focussing on the computation of when/where to produce a reference to a landmark retrieved from the map. From that we will move towards runtime-generated references to landmarks detected in the surroundings. A nice-to-have feature would be references based on an episodic memory, which would compare situations across sessions.

For incidental encounters we have identified several basic types of situations that arise when a robot moves through the clinic, defined a range of possible interaction behaviours and carried out initial WOZ experiments, cf. [10] for more details. The first prototype we are currently developing again uses template-based generation. It remains to be seen whether we will move on to grammar-based generation for these cases.

The object-handing use case is currently very simple, just a few commands and canned system output and/or sounds.

The indoor navigation use case was designed based on actual stated need of the clinic personnel to have robots accompany patients to their accommodation and treatment locations. The incidental encounters are a reality when deploying robots in public spaces. Also the object-handing use case is based on the needs of the clinic personnel and clients. In other words, these use cases were

Figure 2: Robotic platforms used in the INTUITIV project

[1] https://www-cps.hb.dfki.de/research/projects/INTUITIV

Figure 3: Human-robot teamwork in the A-DRZ project

not designed specifically with the aim to do groundbreaking NLG research as such. As stated above, the main research question is how to combine verbal and nonverbal signals to achieve intuitive interaction/communication. The practical challenge is to integrate different components developed by several partners. The main issues we are currently facing have to do with localization in the environment and deciding when exactly to produce verbal output. We need to carry out practical experiments in order to assess the current NLG solution.

Robot-assisted disaster response

Another project I am working on is A-DRZ[2] (Setting up a German Rescue robotics Center), where we address robot-assisted disaster response in complex dynamic outdoor environments, cf. Fig. 3. We are currently developing NLU and teamwork support rather than NLG, cf. [8].

Our observations on human-human interaction provide insights on the communication capabilities robots should ultimately have. First and foremost this includes the generation of *status reports*. These convey (i) *observation reports* from complex dynamic partially damaged/destroyed environments, i.e., descriptions of (potentially unusual) objects and aspects of the environment, which serve the tactical planning of a mission; (ii) *reports of activities and states*, i.e., what the robot has done/seen and plans to do; (iii) *problem or failure reports* conveying what is not going as expected/planned. Status reports should be *interactive*, i.e., essential information is presented first and refined upon request (which may involve further information gathering, i.e., additional perception and interpretation). As for object descriptions, the purpose is typically other than unique identification, e.g., triage status is a priority for victims, hazard assessment for other objects. To support shared situation awareness such interaction is often multimodal, using a graphical/visual interface, such as a map and/or a video feed or sharing static pictures; NLG should be able to support/reflect this. The teams are complex and teamwork involves multiparty communication, so it is necessary to model common ground and NLG should take it into account.

[2] http://rettungsrobotik.de

3 SHARED TASK

Given the above, relevant shared tasks would involve the generation of references to objects and/or activities in the face of uncertainty in a complex dynamic physical situation, co-present or remote, where spoken interaction is combined with other modalities.

This could be for example the *generation of route guidance instructions with references to the actual/current physical environment*. Variations of the task could include the use of different combinations of output modalities (e.g., just speech or speech combined with display; gesture/posture; direction indicators such as arrows or blinkers etc.); robot walking in front, next to or behind the guided person; guiding one person or a group; the environments could include private spaces such as homes (e.g., guiding a quest in an apartment) and public spaces (e.g., office buildings; malls; supermarkets; parking garages; hospitals/clinics; museums, etc.); and could be normal or damaged after an incident/disaster such as an earthquake or an explosion (e.g., consider evacuation-support robots).

4 REFERENCES

[1] Geert-Jan Kruijff, Elke Teich, John Bateman, Ivana Kruijff-Korbayová, Hana Skoumalová, Serge Sharoff, Lena Sokolova, Tony Hartley, Kamy Staykova, and Jiří Hana. 2000. A multilingual system for text generation in three slavic languages. In *Proceedings of the 18th Conference on Computational Linguistics (COLING'2000), July 31 - August 4 2000*, pages 474–480. Universität des Saarlandes, Saarbrücken, Germany.

[2] Elena Karagjosova and Ivana Kruijff-Korbayová. 2002. Conditional Responses in Information Seeking Dialogues. In Kristiina Jokinen and Susan McRoy, editors, *Proceedings of the 3rd SIGdial Workshop on Discourse and Dialogue*, pages 84–87, Philadelphia, Pennsylvania, USA. ACL.

[3] Ivana Kruijff-Korbayová, Tilman Becker, Nate Blaylock, Ciprian Gerstenberger, Michael Kaißer, Peter Poller, Jan Schehl, and Verena Rieser. 2005. Presentation strategies for flexible multimodal interaction with a music player. In *Proceedings of 10th European Workshop on Natural Language Generation (ENLG'05), August 8-10 2005, Aberdeen, Scotland, pages 191–196*.

[4] Tony Belpaeme, Paul Baxter, Robin Read, Rachel Wood, Heriberto Cuayáhuitl, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korbayová, Georgios Athanasopoulos, Valentin Enescu, Rosemarijn Looije, Mark Neerincx, Yiannis Demiris, Raquel Ros-Espinoza, Aryel Beck, Lola Cañamero, Antoinne Hiolle, Matthew Lewis, Ilaria Baroni, Marco Nalin, Piero Cosi, Guilio Paci, Fabio Tesser, Giacomo Sommavilla, and Remi Humbert. Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2):33–53, 2013.

[5] Geert-Jan Kruijff, Ivana Kruijff-Korbayová, John Bateman, and Elke Teich. 2001. Linear order as higher-level decision: Information structure in strategic and tactical generation. In Helmut Horacek, editor, *Proceedings of the 8th European Workshop on Natural Language Generation, Toulouse, France, July 5–6 2001*, pages 74–83.

[6] Ivana Kruijff-Korbayová, Stina Ericsson, Kepa Joseba Rodríguez, and Elena Karagjosova. 2003. Producing contextually appropriate intonation in an information-state based dialogue system. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 227–234. ACL.

[7] Ivana Kruijff-Korbayová; Elettra Oleari, Anahita Bagherzadhal, Francesca Sacchitelli, Bernd Kiefer, Stefania Racioppa, Clara Pozzi. Young Users' Perception of a Social Robot Displaying Familiarity and Eliciting Disclosure. In: Proc. of the 7th International Conference on Social Robotics (ICSR-2015), 7th, October 26-30, Paris, France, Springer, 10/2015.

[8] Christian Willms, Constantin Houy, Jana-Rebecca Rehse, Peter Fettke, Ivana Kruijff-Korbayová. Team communication processing and process analytics for supporting robot-assisted emergency response. In: Proceedings of 2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR). IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR-2019) September 2-4 Würzburg Germany IEEE 9/2019.

[9] Bernd Kiefer, Anna Welker, Christophe Biwer. VOnDA: A Framework for Ontology-Based Dialogue Management. Proceedings of the International Workshop on Spoken Dialogue Systems Technology (IWSDS). Springer, April 2019. p. 12.

[10] Ivana Kruijff-Korbayová, Johannes Hackbarth, Caspar Jacob, Bernd Kiefer, Matthias Schmitt, Tanja Schneeberger, Tim Schwartz, Hanns-Peter Horn and Karsten Bohlmann. Towards Intuitive Verbal and Non-Verbal Communication for Incidental Robot-Human Encounters in Clinic Hallways. In: 15th Annual ACM/IEEE International Conference on Human Robot Interaction, Workshop *The Forgotten in HRI: Incidental Encounters with Robots in Public Spaces*, March 23?26, 2020, Cambridge,

UK. ACM, New York, NY, USA. 2020.