Planning and Realizing Questions in Situated Human-Robot Interaction*

Ivana Kruijff-Korbayová, Geert-Jan Kruijff, Miroslav Janíček
German Research Center for Artificial Intelligence (DFKI GmbH)
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
{ivana.kruijff, gj, mjanicek}@dfki.de

Introduction

This paper is about generating questions in human-robot interaction. Robots need to deal with uncertain or incomplete input about the environment. They can thus benefit from interacting with humans to *acquire information*. Achieving *transparency* is crucial for this to succeed (Thomaz 2006). Robots and humans understand the world differently. When asking a question, a robot needs to say or otherwise signal enough about its belief state and intentions, for the human to clearly understand what it is after.

> *The central thesis we explore here is that transparency of question content, which finds its expression in varying surface form and intonation, can be explained from the combination of social commitment and intentionality, and how these are grounded in a robot's situation awareness.*

In the first part of the paper we survey existing work on the forms and meanings of questions in English. We concentrate on the issue of how besides eliciting information from the hearer, a question can simultaneously offer a window into the speaker's belief state. We discuss the pragmatic effects that result from an interplay between the choice of syntactic form and intonation. In the second part of the paper we propose a formalization based on a notion of common ground, set in a model of situated dialogue as part of collaborative activity (Kruijff et al. 2010a). Explicit modeling of the beliefs and intentions of both the robot and the human enables us to capture the observations from the literature. In the third part of the paper we address the process of generating questions, starting from agent belief modeling through forming the intention to request "missing" information or elicit feedback on "uncertain" information from a human, to planning and constructing the surface realization, including syntax and intonation.

*The research reported here was financed by the EU ICT Programme, Cognitive Systems Unit, projects "CogX: Cognitive Systems that Self-Understand and Self-Extend" (#215181), "Aliz-E: Adaptive Strategies for Long-Term Social Interaction" (# 248116) and "NIFTi: Natural Human-Robot Cooperation in Dynamic Environments" (#247870)

Question Forms and Meanings

Bartels (1999) defines questions as utterances that convey perceived lack of information or speaker uncertainty, regarding a relevant aspect of propositional content. However, uncertainty is not a sufficient condition for a question. A speaker could, for example, just utter a statement asserting their uncertainty. Inspired by Šafářová Nilsenová (2006) we add the speaker's intention to elicit a response as a necessary condition for questionhood. Usually, the speaker of a question intends to elicit a resolving answer that alleviates the uncertainty expressed by the question. However, a speaker may use a question just to raise an issue, irrespective of whether they consider the addressee capable of resolving it.

Research in semantics and pragmatics, e.g., (Pierrehumbert and Hirschberg 1990), (Bartels 1999), (Gunlogson 2001), (Šafářová Nilsenová 2006) identifies differences in the meaning of a question depending on the way it is formulated, in terms of its syntax (interrogative/declarative) and intonation (rising/falling) –cf. the examples below:[1]

(1) What room is this ↓/↑

(2) Is this room a kitchen ↓/↑

(3) This room is a kitchen ↓/↑

In a nutshell, Bartels (1999) argues that a falling phrasal tone expresses assertiveness, i.e., an instruction to the hearer to collaborate actively in the addition of the proposition to the common ground. Utterances with a rising phrasal tone are not assertive in this sense. In a complementary proposal, Gunlogson (2001) submits that falling declaratives commit the speaker to the proposition expressed, while rising declaratives commit the addressee. For polar questions Gunlogson (2001) defines two necessary conditions: uniformativeness with respect to the hearer's commitment set and contingency of speaker's commitment upon that of the hearer. On the other hand, Šafářová Nilsenová (2006) argues that instead the final rise contour corresponds to *epistemic uncertainty*. We try to show how these views can be reconciled.

Formalizing these aspects of question meaning enables us to provide an account of how a robot should phrase its questions.

[1] "↓" stands for final fall, "↑" (Bartels 1999), (Gunlogson 2001).

Formal Framework

We base the formalization on a notion of common ground. This notion is set in a model of situated dialogue, as part of collaborative activity (Kruijff et al. 2010a). The model represents the beliefs and intentions of all robot- and human agents involved in the dialogue. For the robot, these multi-agent beliefs and intentions are spatiotemporally grounded in its perception and knowledge of the environment, i.e. its situation awareness. The model enables a robot to identify, circumscribe, and discuss knowledge gaps (Lison et al. 2010). This allows us to treat questions as a subclass of information-gathering actions in a collaborative activity (Kruijff et al. 2008), providing for a smooth integration with other decision-making processes in a cognitive architecture (Wyatt et al. 2010).

We model a belief as a unit of information about an entity or event referent in reality. We express the information as a probability distribution over alternative interpretations (Lison et al. 2010). Each interpretation is a logical formula, representing an ontologically sorted, relational structure which provides a convenient interface between linguistic meaning, and extra-linguistic forms of meaning representation (Kruijff et al. 2010b). Beliefs are constrained both spatio-temporally and epistemically. They include a frame stating where and when the described referent is assumed to exist, and an epistemic status stating for which agent(s) the information in the belief holds.

We use these belief models to form and maintain common ground in situated, task-oriented dialogue. We use an approach to dialogue processing that follows an intentional perspective Stone and Thomason (2003), Stone (2004), looking at why something is being said (intention), what that something is about (intension), and how that helps to direct our focus (attention). Core to the approach is abductive reasoning. This type of reasoning tries to find the best explanation for observations. In our case, it tries to find the best explanation for why something was said (understanding), or how an intention best could be achieved communicatively (generation). Thereby abduction directly works off the situated, multi-agent belief models the robot maintains. The resulting explanations are defeasible: The robot acts upon them, but if they turn out to be wrong, the robot can revise the explanation and the beliefs it was based on (e.g. through further interaction) and thus adjust its common ground.

We propose an interpretation of *social commitments* in terms of such multi-agent beliefs. This allows us to explicitly reason with them, promote them to common ground, and consequently draw conclusions about the expected future progression of the interaction. Social commitments capture a part of the social aspect of interaction. Such a commitment is a public (or so perceived) state oriented towards the social group (human, robot), committing the interlocutors to certain rules of behaviour. In the simplest case, we can consider behaviour rules of the form

$$\textit{trigger} \rightarrow \textit{future-effect}$$

expressed as pairs of beliefs. This sets our approach apart from approaches that model social commitment as an irreducible construct (**?**).

Context, Intentions and Commitments

Based on the notion of multi-agent belief state that includes common ground, we introduce intentions as precondition-postcondition pairs on the state. In general, an intention is a goal that an agent is committed to achieving[2]. The agent should refrain from acting in such a way that renders the intention (goal) unachievable (cf. **?**).

Preconditions are applied to the state in which the corresponding intention can be realized. Postconditions specify the conditions that must hold in the resulting state after realizing the intention. In a sense, postconditions specify the sufficient conditions to consider the intention fulfilled, and preconditions specify the necessary conditions.

An intention to fill in a knowledge gap can be realized in several forms. Since the form of a question also influences the expected answer, it is obvious that it is not enough just to specify a question as a function of the knowledge gap to be filled, but it is necessary to consider the form of the question in the decision-making process.

We argue that the form can be inferred from the social commitments the question should appeal to in a relatively straightforward manner. In terms of commitments, we can analyze the examples in (1)-(3) as follows:

- In (1), the robot makes the claim that the user is responsible for filling in the gap. This responsibility is based on the robot's beliefs about the human's knowledge or knowledgeability. This in turn is inferred from the robot's beliefs about the interpersonal aspect of the interaction (roles).
- In (2), the robot proposes a single hypothesis and holds the user responsible for defending ("yes") or refuting ("no", "this is a living room") this hypothesis. Note that it might not be the *best* hypothesis—but merely the hypothesis that is most worthy of verification (e.g. based on overall utility). For example: in a search-and-rescue, the robot might ask "is this a person?" because getting an answer greatly influences the future course of action. So a "yes" here might trigger a change in the robot's behaviour (e.g. a switch toward making sure that the scene is safe for human rescuers, similarly, a "no" would mean that the robot doesn't have to pay (that much) attention to the object in question any more, and can carry on exploring the area. In other words, even if the robot has a better hypothesis about what the object in question is, it might be *rational* to get the possibility that it is a person off the table.
- In (3), the robot expresses his commitment to the claim: it commits to the defence of the claim. Should the user ask him "why?", the robot should be able to reveal its justification for the beliefs.

In our model, we treat these social commitments as beliefs in the *preconditions* of the intentions to ask. For instance, in (1), there are at least two distinct preconditions that allow the realization of the appropriate question:

(a) the robot believes that the human knows what room this is;

[2]Note that this commitment is *not* to be conflated with the social commitment mentioned above. See for instance **?**

(b) it is common ground that the human will defend the claim that he knows what room this is.

The intention is created as an interpretation of an abductive proof to achieve the desired effect (the robot's knowledge of the room type), given current context. As a product of abduction, it is defeasible. For instance, should it turn out that (b) does not hold, this particular proof will be retracted.

Planning and Realization of a Question

The preceding sections show how a question comes about. The robot continuously maintains a model of the environment and agents acting therein, based on its observations and interaction with others. Within this model, knowledge gaps can be actively identified, and trigger a need to address them through interaction (Kruijff et al. (2008), (Wyatt et al. 2010). The approach we present here then abductively infers an intentional structure which is grounded in this belief model, and indicates (applicable) social commitments concerning the expected continuation of the dialogue after the question has been posed.

The information contained in this intentional structure is sufficient to account for the kinds of variations in the form and meaning of questions typically observed. Content planning uses the structure to make decisions about functional content structure of the question to be generated, including information structure status of individual referents. Content planning yields a fully specified logical form which we can then realize as a surface string with intonational markup in OpenCCG (White and Baldridge 2003), (Kruijff et al. 2010b), (Kruijff- Korbayova et al. 2011), and synthesize with the Mary text-to-speech system (Schröder and Trouvain (2003).

References

C. Bartels. *The intonation of English statements and questions: a compositional interpretation.* Routlege, 1999.

M. Bratman. Intentions, Plans, and Practical Reason. Harvard University Press, Cambridge, MA, USA, 1987.

J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. van der Torre. The boid architecture: conflicts between beliefs, obligations, intentions and desires. In *Proceedings of the fifth international conference on Autonomous agents,* AGENTS 01, pages 916, New York, NY, USA, 2001. ACM.

C. Castelfranchi. Modelling social action for AI agents. *Artificial Intelligence,* 103: pp. 157–182, 1998.

C. Gunlogson. *True to Form: Rising and Falling Declaratives as Questions in English.* PhD thesis, University of California at Santa Cruz, 2001.

G. J. M. Kruijff, M. Brenner, and N. A. Hawes. Continual planning for cross-modal situated clarification in human-robot interaction. In *Proceedings of the 17th International Symposium on Robot and Human Interactive Communication* (RO-MAN 2008), 2008.

G. J. M. Kruijff, M. Janíček, and P. Lison. Continual processing of situated dialogue in human-robot collaborative activities. In *Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication,* 2010a.

G. J. M. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, H. Zender, and I. Kruijff-Korbayová. Situated dialogue processing for human-robot interaction. In H. I. Christensen; G. J. M. Kruijff; J. L. Wyatt, editor, *Cognitive Systems,* volume 8 of *Cognitive Systems Monographs, COSMOS,* chapter 8, pages 311– 364. Springer Verlag, Berlin/Heidelberg, Germany, 2010b.

I. Kruijff-Korbayová, R. Meena, and P. Pyykönen. Perception of visual context and intonation patterns in robot utterances. In *Proceedings of 6th ACM/IEEE International Conference on Human-Robot Interaction* (HRI 2011), 2011.

P. Lison, C. Ehrler, and G. J. M. Kruijff. Belief modelling for situation awareness in human-robot interaction. In *Proceedings of the 19th International Symposium on Robot and Human Interactive Communication* (RO-MAN 2010), 2010.

J. Pierrehumbert and J. Hirschberg. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication,* pages 271–312. MIT Press, Cambridge, MA, 1990.

M. Schröder and J.Trouvain.The german text-to-speech synthesis system mary: A tool for research, development and teaching. *International Journal of Speech Technology,* 6:365377, 2003.

M. Stone. Intention, interpretation and the computational structure of language. *Cognitive Science,* 28 (5): 781809, 2004.

M. Stone and R. H. Thomason. Coordinating understanding and generation in an abductive approach to interpretation. In *Proceedings of DIABRUCK 2003: 7th workshop on the semantics and pragmatics of dialogue,* 2003.

A. L. Thomaz. *Socially Guided Machine Learning.* PhD thesis, MIT, June 2006.

M. Šafářová Nilsenová. *Rises and Falls: Studies in the semantics and pragmatics of intonation.* PhD thesis, Institute for Logic, Language and Information, Universiteit van Amsterdam, 2006.

M. White and J. Baldridge. Adapting chart realization to CCG. In *Proceedings of the Ninth European Workshop on Natural Language Generation,* Budapest, Hungary, 2003.

J. L. Wyatt, A. Aydemir, M. Brenner, M. Hanhiede, N. Hawes, P. Jensfelt, M. Kristan, G. J. M. Kruijff, P. Lison, A. Pronobis, K. Sjöö, D. Skočaj, and A. Vrečko. Self-understanding and self-extension: a systems and representational approach. *IEEE Transactions on Autonomous Mental Development,* 2(4): 282303, 2010.