EU FP7 CogX
ICT-215181
May 1 2008 (52months)

DR 6.5: Mixed Initiative Situated Dialogue-Guided Curiosity

Hendrik Zender, Miroslav Janíček, Geert-Jan M. Kruijff

DFKI GmbH, Saarbrücken
⟨zender@dfki.de⟩

Due date of deliverable:	Month 48
Actual submission date:	May 28 2012
Lead partner:	DFKI
Revision:	final
Dissemination level:	PU

In previous years, WP6 investigated how situated dialogue could be used in human-robot interaction to help the robot learn more about its environment. This involved grounding dialogue in multi-agent models of beliefs and intentions, dealing with the uncertainty and incompleteness in these models, and communicating about the content in these models at different levels of granularity. These dialogues were typically tutor-driven. In Year 4, WP6 explored topics that have to do with robot-initiated dialogues. We investigated issues in common ground and transparency that help a robot to make use of its dialogue capabilities to explain its internal state and past actions to its user as well as to learn about the world by asking for missing knowledge or for clarifying uncertain knowledge. Furthermore, Year 4 was used to consolidate longer-term efforts originating in previous project periods, such as a software toolkit for natural language dialogue processing for talking robots.

Executive Summary

One of the objectives of CogX is self-extension. This requires the robot to have an intrinsic motivation – *"curiosity"* – to gather information, in order to acquire new and revise old knowledge. One of the sources of information about the world is dialogue. For dialogue to work, the robot needs to be able to establish with a human some form of mutually agreed-upon understanding, a *common ground*. This requires the robot to provide *transparency* about its actions and goals, as well as *interpreting the intentions* of its interlocutor. The overall goal of WP6 is to develop adaptive mechanisms for situated dialogue processing, to enable a robot to establish such common ground in situated dialogue.

In Year 1, WP6 investigated how a robot could carry out a situated dialogue with a human, about items in the world it needed to learn more about. The robot was able to formulate questions against a multi-agent model of situated beliefs, indicating what it did and did not know – and what it would like to know. The robot was able to represent and reason with uncertainty in experience, but it was relatively fixed in the strategies it would follow to communicate with the human about resolving the uncertainty.

In Year 2, WP6 investigated several issues in how to make dialogue behavior more *adaptive*. This covered several aspects: (1) Making dialogue strategies more adaptive, and (2) varying how much a robot needs to describe to be optimally transparent.

Throughout Years 1 and 2 we assumed the robot to have a fixed set of communicative competences, particularly where it concerned grammatical resources. Practically this meant that, even though the robot was still learning more and more about the world, it already knew how to talk about it. In Year 3, WP6 shifted its focus on taking the CogX objective of self-extension to the realm of situated dialogue processing as well. Making use of large ontologies and on-line resources for modeling common sense indoor knowledge (OMICS, WordNet, Bing image search) provided a powerful means for large-coverage resources for communicating about indoor environments. At the same time, we broadened the scope of self-extension in situated dialogue to the aspect of language acquisition per se.

In Year 4, WP6 investigated issues in *common ground* and *transparency*. We focused on verbalizing a robot's internal state and background knowledge, and integrating that functionality into human-robot interaction. This helps the robot to *explain* its internal state and past actions to its user, and to *learn* about the world by asking for missing knowledge or for *clarifying* uncertain knowledge. The result is a context-adaptive approach for clarification and explanation (Task 6.7), based in a robot's own understanding of what it does (not) know, and what it can (or cannot) attribute to a human it interacts with. The robot can initiate these kinds of dialogue itself, motivated by its own curiosity (Milestone M6.5). Over and beyond these efforts,

we consolidated our longer-term efforts originating in previous project periods, resulting in a software toolkit for natural language dialogue processing for talking robots (TaRoT).

Overall, the work in WP6 has led to a comprehensive theoretical and practical framework for situated dialogue, in which we have paid particular attention to issues in self-extension and introspection in and through dialogue. Our story of the situated nature of spoken dialogue in human-robot interaction is based in a situationally, socially, and epistemically perspectivized (i.e. "asymmetric") notion of common ground (Yr4) and acquisition (Yr3), and a cross-modal view on the relation between linguistic and extra-linguistic information. We consider situated dialogue in the larger context of a collaborative activity, using common ground and cross-modal connections between information to establish intention, intension and denotation for utterances – both for comprehension, and production, and for entities it does or does not know about. The robot can verbalize and clarify this knowledge, in context-adaptive ways, to establish transparency in common ground with the user (introspection), and to drive deliberated, self-motivated forms of learning (self-extension).

Role of Situated Dialogue in CogX

CogX investigates cognitive systems that self-understand and self-extend. In most of the scenarios explored within CogX such self-extension is done in a mixed-initiative, interactive fashion (e.g. the George and Dora scenarios): the robot interacts with a human, to learn more about the environment. WP6 contributes situated dialogue-based mechanisms to facilitate such interactive learning. Furthermore, WP6 explores several issues around the problems of self-understanding and self-extension in the context of dialogue processing. The dialogue capabilities provided by WP6 enable the robot to conduct situated dialogues for interactive learning based on its own curiosity. The models for spatially situated dialogue are grounded in the spatial models developed in WP3 (see also DR.3.3 and DR.3.4, especially [41]). The approach to continual abductive dialogue interpretation has a strong relation to WP4, in that it is inspired by the continual planning approach taken in that WP. At the same time, verbalization of past, ongoing, and past actions is grounded in the planning representations developed in WP4.

Contribution to the CogX Scenarios and Prototypes

The work of WP6 presented in this deliverable, DR.6.5, contributes directly to the George and Dora scenarios, in relation to work performed in WP1 (generating motivation goals from dialogue), WP3 (Qualitative spatial cognition), WP4 (Planning of action, sensing and learning), WP5 (Interactive continuous learning of cross-modal concepts), and WP7 (Scenario-based integration). In particular, DR.7.3 reflects the work from WP6 in the systematic use of dialogue as a means of knowledge gathering and clarification, and for establishing transparency. DR.7.4 (especially [31]) describes a robot that is capable of continuous learning of visual concepts in dialogue with a tutor. These learning dialogues can be initiated by the human tutor or by the system itself.

1 Tasks, Objectives, Results

1.1 Planned work

The overall goal of CogX is to arrive at a theory of cognitive robots which are capable of self-understanding and self-extension. During the last years, WP6 worked on adaptive mechanisms for situated dialogue processing that would enable a robot to discuss with a human what it did and did not understand about the world. And, thus, through such dialogue, it could gain information to help it learn more. While in the previous years, such dialogues were mainly tutor-driven, the focus in Year 4 is on robot-initiated dialogues. The planned work for WP6 in Year 4 is to support curiosity-driven self-extension through situated dialogue and to provide means for making the robot's self-understanding transparent by verbalizing explanations of its internal states, its decision-making and actions.

Task 6.7 ***Adaptive strategies for clarification and explanation.*** *Towards the end of the project, the robot's learning is primarily curiosity-driven. This is an advance in that it now actively needs to initiate dialogues, if it wants to interact with other agents. We therefore want to investigate (adaptive strategies for) clarification and explanation, more from the engagement-level [30], to address the issue of how to set the context for a clarification request (i.e. scaffolding it), to avoid "out-of-the-blue" behavior.*

Milestone M.6.5 ***Mixed initiative situated dialogue-guided curiosity.*** *The system will be able to initiate and drive situated dialogues for interactive learning based on its own curiosity.*

Objective 2 *Specific representations of beliefs about beliefs for the specific cases of dialogue, manipulation, maps, mobility and some types of vision.* [WPs 2,3,6]

Objective 3 *Representations of how actions will alter the belief state of the cognitive system, and those of other agents, as represented in the first two objectives, i.e. models of the effects of actions on beliefs about space, categorical knowledge, action effects, dialogue moves etc.* [WPs 1,2,3,4,5,6]

Objective 9 *Methods that enable a robot to represent and reason about its beliefs and those of other agents to support natural dialogue and to extend its own abilities and understanding.* [WP 6]

Objective 11 *A robotic implementation of our theory able to complete a task involving mobility, interaction and manipulation, in the face of novelty, uncertainty, partial task specification, and incomplete knowledge.* [WPs 2,3,6,7]

1.2 Actual work performed

In Year 4 we developed new approaches for verbalizing a robot's internal belief state, reflecting what it does and does not know, and what it has done so far. We have integrated this type of verbalization into our general approach to situated dialogue processing. This involves tying it in with our algorithms for situated referring expression generation and resolution, and anchor progression. The primary function of this verbalization is to make the robot's understanding of, and reasoning about, the world more transparent to the user (Task 6.5). This aids building up and maintaining common ground in human-robot interaction, which is crucial if the robot is to successfully use dialogue to obtain information from a human user (Milestone M6.5).

We have achieved this task, and the corresponding milestone, as follows:

1. Building on earlier WP6 work we have further developed our abduction-based approach to continual dialogue understanding (§1.2.1). This makes it possible to reason explicitly with knowledge gaps, be they attributable to the human or the robot itself, and their possible resolution through dialogue as a form of action within a collaborative activity.

2. The very fact that we have knowledge gaps, and the obvious observation that robots and humans experience reality differently, have led us to reconsider the notion of common ground (§1.2.2). Already in earlier work on our abductive model we dropped the assumption on symmetry between dialogue participants. We have now taken this to its logical consequences, and built up a model of common ground which starts from asymmetry in the social, situated, and epistemic factors which play a role in common ground.

3. Finally, it is against this "asymmetric" background that we then consider verbalization of internal state (§1.2.3). This state covers robot beliefs, and past actions leading up those beliefs. The function of such verbalization in dialogue is to explain what the robot believes, and why it believes what it believes. This transparency, realized as a sub-dialogue clarifying what the robot would like to know and why, then provides the scaffolding for whatever question the robot is motivated to ask (mixed-initiative dialogue).

A (partly stylized) example of such a sub-dialogue is as follows.

1. I moved from the lab to the kitchen,

2. to look for cornflakes

3. I have searched for the cornflakes,
4. but I am unable to find them.
5. Could you tell me where the cornflakes are?

The example illustrates the verbalization of different plan/action steps (1,3), reasons (2), and outcomes (4). This provides a background for formulating the final question (5). Annex 1.2.3 describes in detail how the sub-dialogue in (1)–(4) can be generated from a plan and its execution trace.

With this we also answer the reviewers comments, pertaining to WP6:

> *[T]he methods for interpreting and producing referring expressions still need some work and are not yet a particularly convincing showcase of the very interesting model of situated dialogue being developed in the project.*

The work performed in Year 4 in WP6 (and through the integration in WP7) brings back many of the different strands worked on over the years. Using the integrated system functionality, we can now showcase our approach in various complex settings in human-robot interaction for self-extension.

The work performed in this WP meets several main objectives of CogX. The approaches to abductive dialogue interpretation (§1.2.1) and modeling common ground (§1.2.2) make contributions to Objectives 2, 3, 9 by providing methods for representing and reasoning about the beliefs of the robotic agent and other (human) agents it interacts with. The work on situated plan and execution verbalization (§1.2.3) makes further contributions to objectives 3 and 9 by allowing the robotic agent to verbalize its past actions and intentions, thereby allowing a human agent to understand the robot's behavior and beliefs. Transparency of a robot's intentions and actions has been further investigated in the context of a special journal issue (§1.3.2). Additionally, the work on dialogue interpretation and verbalization has been implemented in a software toolkit for talking robots (§1.3.1), which has been deployed in the WP7 integrated systems (cf. Objective 11).

1.2.1 Abductive dialogue interpretation

In task-oriented dialogues between two agents, such as between two humans or a human and a robot, there is more to dialogue than just understanding words. An agent needs to understand what is being talked about, and it needs to understand why it was told something. In other words, what does the speaker *intends* the hearer to do with the information, in the larger context of their joint activity? Language understanding can thus be phrased as an *intention recognition* problem: given an utterance from the human, how do we find the intention behind it?

Abductive Reasoning for Continual Dialogue Understanding

Janíček (Annex 2.1) presents an extended model of the abductive continual approach to situated dialogue understanding. This model draws inspiration from the field of continual planning [6], by explicitly capturing the possible knowledge gaps in such an interpretation. The idea is based on the notion of assertion*, an explicit test for the validity of a certain fact, going beyond the current context. This makes it possible to deal with both uncertainty and incompleteness in situated dialogue processing.*

Let us briefly discuss an example that uses this mechanism. A more detailed example can be found in the article in Annex 2.1.

Suppose that a human user is dealing with a household robot capable of manipulating objects (picking them up, putting them down). The robot and the human are both looking at a table with a mug ("mug_1"), and the human wants the robot to pick up the mug. The human's utterance, "Take the mug" is first parsed and analyzed semantically, and its translation is made part of the abduction context c, within which the robot tries to make sense of the utterance. The inference establishes several alternative proofs, and weighs them by the "costs" (probabilities) for the individual facts and assumptions appearing in a proof. The best proof is the one with the lowest cost. Suppose that the best proof state returned by ABDUCE is the following:

$\text{uttered}(\text{human}, \text{robot}, \text{event}_1)$	$[\textit{proved}]$	(1)
$\text{proposition}(\text{event}_1, \text{take})$	$[\textit{proved}]$	(2)
$\text{intends}(\text{event}_1, \text{human}, I)$	$[\textit{assumed}(p = 0.9)]$	(3)
$\text{relation}(\text{event}_1, \text{patient}, \text{thing}_1)$	$[\textit{proved}]$	(4)
$\text{refers\_to}(\text{thing}_1, X)$	$[\textit{asserted}]$	(5)
$\text{pre-condition}(I, \text{object}(X))$	$[\textit{asserted}]$	(6)
$\text{post-condition}(I, \text{state}(\text{is-holding}(\text{robot}, X)))$	$[\textit{assumed}(p = 0.7)]$	(7)

The proof is an explanation of the event in terms of a partially specified intention I related to the task specified above. An explanation is defined by its pre- and post-condition. The precondition is the existence of an entity X, and the postcondition is the state in which the robot is holding the entity X.

Assumptions are made with an assumability probability according to the beliefs the robot currently maintains, and inferences it can make from more general background knowledge (i.e. a rule base). In our approach the assumability function is manually designed, but it is conceivable to learn or infer it automatically. Reference resolution (i.e., does "the mug" refer to mug_1 or to some other referent?) is therefore essentially treated as an abduction problem.

Note that the proof state contains two atoms marked as assertions. These are the explicit gaps in the proof that make it a *partial interpretation.* They are chosen by the domain engineer, and since they need to be verified (or falsified) by an external process, they form the interface to external knowledge bases and decision-making, which will select some of the assertions, and tries to verify them.

Suppose that the assertion (5) is tested first. This amounts to resolving the referring expression represented by $thing_1$. Under the open-world assumption, two interpretations are conceivable: the referent is mug_1 or the human might be referring to an object that is not part of common ground, and the reference thus cannot be resolved. The commitment to one of these interpretations is made by taking into account the probabilities and uncertainties about the world that are represented in the robot's beliefs.

The next assertion (6) expressed the presupposition accommodation that there exists an object to which the human is referring. This opens the possibility for further clarification in case the reference in (5) could not be resolved with a sufficient level of confidence.

In case the reference has been resolved to mug_1, but with low confidence, the robot might ask "Did you mean I should take *this* object?" (pointing at the mug, testing the hypothesis $\text{pre}(I, \text{object}, \text{mug}_1)$).

Likewise, in case the robot abductively concludes that the human is referring to an unknown mug, it might ask "Which object did you mean?", prompting the human to give an answer that would ultimately become the proof of the test action for $\text{pre}(I, \text{object}(X))$.

Alternatively, the robot might simply bring the most likely object. The human's acceptance of the choice would then verify the assertion. This is, again, a matter for consideration in the external planning and goal management.

1.2.2 Common ground

In order to interpret what it is that is being communicated, one needs to construct a *meaning representation*. In the processes of constructing meaning, one can appeal to different sources of information. In situated dialogue, these sources at least encompass the situation being described (*focus situation*); any other physical or discursive contexts, or common knowledge (*resource situations*); and the ways in which the communication partners take part in the dialogue (*social situation*) [11]. Figure 1 provides an example.

The social situation in situated dialogue makes it clear that communication partners look at the world from different perspectives. Some of the effort in communicating therefore goes into establishing a *common ground* between partners. The point is to establish a mutual understanding of what is being talked about, what is appealed to notably in reference to the world [8, 7]. This is a dynamic process, in which partners coordinate and align their beliefs [39, 28].

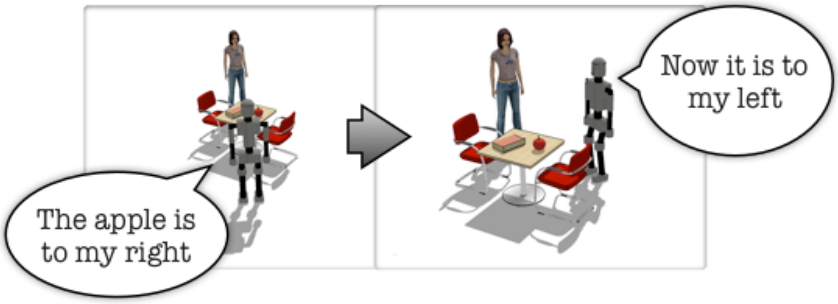

Figure 1: Different situations in meaning: "Now it is to my right" appeals in contrast to the previous situation ("now"), preceding dialogue ("it"), the current situation ("is"), and the current social situation ("to my right").

This need for common ground is not restricted to communication between humans. It holds just as much for human-robot interaction, where understanding and facilitating different perspectives on the world is crucial to establishing an effective collaborative context.

There is no Common Ground in Human-Robot Interaction

There is an inherent asymmetry to situated communication, especially in human-robot interaction. Robots literally see the world differently. This raises an important issue for how to model common ground between a human and a robot. Traditional approaches to common ground adopt a possible worlds-like model theory, on which a shared belief represents the fusion of two private beliefs within a single model. This silently assumes a symmetry in categories against which both private beliefs can be interpreted: A symmetry which cannot be assumed for human-robot interaction. Hence, on such a model, there is no common ground between humans and robots. Kruijff (Annex 2.2) presents an alternative model theory, which captures the inherently subjective nature of experience. It is based on a notion of propositions-as-proofs, turning subjective interpretation as well as the projection to intersubjective verification into a notion of inference or argumentation. Beliefs are arguments, whether private, attributed or shared. This results in a formulation of common ground as a dynamic structure of always argued, but possibly only partially confirmed, or partially assumed, beliefs.

Common ground is a complex notion. Generally speaking, it is about something *between* communicative partners: Namely, "shared understanding." Minimally, this is a shared understanding of the entities which have been talked about, grounded in the (possibly situated) domain of discourse [8, 7, 39]. After entities have been introduced into common ground, they are accessible for reference. Naturally, there can be more to common ground. In task-oriented domains, common ground provides an interface between the task domain itself, and the communication which mediates and coordinates collaboration in that domain. Typically, this then leads to considering intentions, plans, and tasks to be part of common ground as well. On top of which we can essentially consider all that is implied by the concept of situated meaning [11], illustrated e.g. in Figure 1. (And that would include "commonsense knowledge," e.g. what if the robot was discussing not an edible apple, but a Mac?) As said, common ground is a complex notion.

1.2.3 Verbalization of plans

Interactive intelligent robots need to possess two important features: *autonomy* and *communication skills*. Situated communication can comprise different modalities, like, e.g., spoken dialogue. Autonomy can range from simple reactive control loops to sophisticated goal-directed action planning and execution. If the robot and its user are supposed to engage in some form of collaborative activity – no matter if it is a cognitive assistant whose main purpose is to support its human user, or if it is a curious robot whose main drive is to learn about the world – the human crucially needs to be able to know and understand the robot's actions (past, present, as well as intended ones) in as far as they are relevant for achieving a task at hand; the robot's autonomous decision-making and acting need to be *transparent* to the human.

We consider the verbalization of plans as one way of achieving such a *transparency* in situated discourse. In the context of CogX, we want to endow our robots with the ability to tell a human user about present objectives, what actions were executed in the past and what further actions are planned.

Situated Plan and Execution Verbalization

We present an approach to verbalizing reports of intentions and actions of a planner-enabled agent. We consider the case of an interactive intelligent robot that is endowed with a symbolic AI planner. The robot uses the planner to determine and execute sequences of actions in order to achieve a given goal. Since the robot is operating in a real, physical environment, making use of possibly imperfect sensing and actuating, it is likely to encounter unforeseen events or failures, and therefore needs to re-plan in order to come up with alternative plans for achieving its goal. In order to tell a human about what was planned, what was and was not successfully done, what happened – and why – we present a method for generating natural language reports based on such intended plans and the event structure of their execution. This verbalization is inherently situated in nature. For one, external entities that are used in the planning process refer to things, persons, or locations in *the physical environment that the robot and the hearer are situated in. Secondly, the robot's intentions and actions are temporally related to the discourse situation: the robot might report about its current plans and attempts for executing an ongoing task; the robot might explain what it did and what it couldn't successfully do in order to establish common ground for asking the human for help; or the the robot might simply report about past events and give details about why it chose to do what.*

The approach has been implemented on the integrated robotic system Dora [17]. Dora is programmed to exhibit a variety of intelligent behaviors, among which other intelligent mechanisms might arbitrate. However, these decisions might (initially) be intransparent to a human observer. We hence want Dora to provide verbal reports of what it attempted to do and why – thus establishing transparency about the complex spatio-temporal and causal relationships of its actions and action attempts.

Dora is equipped with a switching symbolic-probabilistic AI planner [12], able of continual planning and re-planning [6]. The approach is integrated in the general natural language processing sub-system of Dora and shares its linguistic resources with the other dialogue capabilities [19]. The verbalization module is connected to Dora's spatial knowledge base in order to refer to entities in Dora's spatial environment [40, 41, 42].

1.3 Additional work performed

1.3.1 Software toolkit for situated dialogue processing

In CogX, we gathered substantial knowledge and experience in developing functionality for situated dialogue processing. This extended the experience we had already gained in the project preceding CogX, namely CoSy. In Year 4, we dedicated effort to consolidating this knowledge and experience in a toolkit. This toolkit, called the Talking Robots Toolkit or TAROT provides a set of reusable functionalities and resources to build dialogue systems for human-robot interaction.

TAROT– The Talking Robots Toolkit

TAROT *is an open-source software framework for building spoken dialogue functionality for human-robot interaction.* TAROT *does not impose a specific (cognitive) architecture for building a dialogue system. Its framework allows for multi-threaded (or asynchronous) processing. Processes are defined as glass boxes [25] (not black-boxes) and interact in an event-driven fashion.* TAROT *is written in the Scala programming language and targets the Java platform. Annex 2.4 provides a technical description.*

1.3.2 Outreach: state-of-the-art in expectations, intentions, and actions in human-robot-interaction

Human-robot interaction as a scientific field has received much attention in the past years. The research performed in the context of CogX, especially concerning situated natural language processing and human-robot spoken communication, draws from and directly contributes to this field.

One tangible and relevant outcome of this was that researchers from the CogX consortium (Marc Hanheide, Geert-Jan M. Kruijff, Hendrik Zender) organized the ICRA 2010 Workshop on Interactive Communication for Autonomous Intelligent Robots (ICAIR). Its topics centered around making robots articulate what they understand, intend, and do. Being a successor to the ICRA 2008 Workshop on Social interaction with Intelligent Indoor Robots (SI3R), it attracted researchers from different fields of robotics who work on robots that communicate.

As a follow up to this workshop as well as to the HRI 2011 Workshop on The Role of Expectations in Intuitive Human-Robot Interaction (Verena Hafner, HU Berlin; Manja Lohse, Bielefeld U; Joachim Meyer, Ben-Gurion University of the Negev, Israel; Yukie Nagai, Osaka U; Britta Wrede, Bielefeld U), the editors of the International Journal of Social Robotics proposed to organize a Special Issue on Expectations, Intentions and Actions, for which Marc Hanheide, Manja Lohse, and Hendrik Zender served as guest editors.

Journal of Social Robotics: "Expectations, Intentions & Actions"
This special issue (Annex 2.5) bundles recent advances in embodied and situated social human-robot interaction. The key questions are how meeting or failing to meet the user's expectations influences the efficiency and effectiveness of human-robot interaction; how more effective and efficient interaction with humans can be achieved using modalities available to a robot; how robots can be equipped with models enabling them to understand their users' state of mind; and similarly, how they can make their own expectations and states explicit through eligible communication channels. Each of the seven contributed articles in this issue highlights different aspects around the central theme of expectations, intentions, and actions in human-robot interaction*. The topics covered range from recognition of verbal and non-verbal cues of intentions and expectations, to verbalization and presentation techniques that make internal processing of the robot accessible to the human.*

1.4 Relation to state-of-the-art

Below we briefly discuss how the obtained results relate to the current state-of-the-art. We refer the reader to the annexes for more in-depth discussions.

Abduction Our approach to situated dialogue processing is based on our ongoing research in dialogue as part of continual ("contingency-based") collaborative activity ([23], and [24]). Dialogue modeling is connected to multi-agent models of situation awareness. These models capture beliefs and intentions, and their inherent uncertainty and incompleteness with respect understanding the environment. It is a strongly intention-oriented approach, in the sense of [2, 1, 14, 29, 9]. It accords a strong role to common ground in interaction [21, 20]. Core to the approach is abductive inference. This is inspired by [34, 35, 33], but differs in that our approach does not assume symmetry in understanding between the different dialogue partners. The recent work reported in this deliverable illustrates how we take asymmetry in understanding to dealing with partial information (Annex 2.1), and to a reconsideration of the subjective and intersubjective nature of content in common ground (Annex 2.2)

Common ground Understanding and facilitating *common ground* in human-robot interaction has received substantial attention in recent years [21, 22, 20]. As put forward by several researchers [32, 38, 36, 27], common gound is indeed crucial for establishing an effective collaborative context. Failure to do so typically leads to a breakdown in communication, see e.g. [36, 29]. There is a large body of work on considering intentions, plans, and tasks to be part of common ground [2, 14, 1, 29, 16, 3, 15, 9, 23]. The problem common to all these approaches is that they assume a symmetry between interlocutors: How the speaker sees and talks about the world is how the hearer understands the world. This symmetry-assumption is explicitly stated in e.g. [35]. We can also see it reflected in the formal aspects of the model theories underlying these approaches. Common ground on a belief in proposition p means that, in a single possible worlds model, we can reach a world on which p holds from the worlds on which the private and attributed beliefs about p hold. This is incorrect, as it assumes that the different interlocutors have a single (symmetric) objective model for interpreting. This ignores the fundamentally *subjective* nature of experience, and the inherent differences between humans and robots. The algebraic model theory we propose here overcomes these problems.

Plan and execution verbalization Brenner [4] describes the use of classical AI planning techniques for interpretation and execution of human commands. He sketches how a robot can understand natural language (NL), plan the realization and revise its plans based on new perceptions. Their approach is similar to ours in that the goal is to couple planning symbols to natural language semantics and surface forms. In contrast to the approach presented in this paper, their approach focusses on *understanding* rather than on *generation*.

However, what we are interested in is the generation of natural language from planned and executed goal-directed sequences of actions. One domain where such an approach is chosen is story telling. Telling stories requires methods from many subfields of artificial intelligence (AI), e.g. planning, reasoning about beliefs, and dialogue systems. In [5] an approach to story generation using a continual multi-agent planner is presented. The Virtual Storyteller is another framework that generates simple story texts [37]. It is based on simulation of virtual characters in a story world. An event sequence is captured by a 'Plot Agent' in a formal representation. The representations are similar to the plans in [5], i.e., STRIPS-like [10]. A 'Narrator' component turns the representation into an actual story by selecting the content and processing it with NLG techniques. Story telling and NLG are also brought together in, e.g., [26].

Another system that verbalizes some kind of plans is PROVERB [18]. In this work, mathematical natural deduction (ND) style proofs are verbalized. As input there is a representation of a ND proof. It is processed by a macro-planner to plan output that consists of primitive actions. The actions can be defined as communicative goals they fulfill as well as their possible verbalizations. Subsequently, more detailed linguistic decisions are made in the micro-planning component. Syntactic realizations are done using Tree-Adjoining Grammar (TAG). The final output then is the ND proof in natural language.

When verbalizing plans and actions for autonomous robots, an additional aspect comes into play: the agent's observations are potentially incorrect or incomplete, and execution (as well as execution failures) become a key issue. In such a context it is desirable to have a formal way to determine an explanation of why a plan went wrong and how the problem could be solved. Göbelbecker et al. [13] provide a formalization regarding this issue. The work presented here addresses the prerequisites for informing the user about failures executing the plan and how the user could help to solve the problems. In order to achieve this, we investigate appropriate strategies for a suitable verbalization of the planned and performed actions in a way that is understandable to a human user.

2 Annexes

2.1 Janíček, "Abductive Reasoning for Continual Dialogue Understanding"

Bibliography Miroslav Janíček. "Abductive Reasoning for Continual Dialogue Understanding." In M. Slavkovik and D. Lassiter, editors, *New Directions in Logic, Language, and Computation.* Springer, 2012 (to appear).

Abstract This paper presents a continual context-sensitive abductive framework for understanding situated spoken natural dialogue. The framework builds up and refines a set of partial defeasible explanations of the spoken input, trying to infer the speaker's intention. These partial explanations are conditioned on the eventual verification of the knowledge gaps they contain. This verification is done by executing test actions, thereby going beyond the initial context. The approach is illustrated by an example set in the context of human-robot interaction.

Relation to WP The paper presents an extended and improved version of the approach presented in DR.6.3. It provides the basic inference mechanism for reasoning about beliefs and intentions in the context of dialogue processing. In the context of DR.6.5 it is instrumental in determining the appropriate epistemic context for scaffolding mixed-initiative dialogue for curiosity-driven learning – i.e. which beliefs held by the robot (private or attributed) need to be verbalized, to explain what the robot does or needs to know.

2.2 Kruijff, "There Is No Common Ground In Human-Robot Interaction"

Bibliography Geert-Jan M. Kruijff. "There Is No Common Ground In Human-Robot Interaction." *Manuscript*, 2012.

Abstract There is an inherent asymmetry to situated communication. Those communicating look at the world from different perspectives. This holds particularly true for human-robot interaction. Robots literally see the world differently: They experience reality in fundamentally different ways. This raises an important issue for how to model common ground between a human and a robot. Traditional approaches to common ground adopt a possible worlds-like model theory, on which a shared belief represents the fusion of two private beliefs within a single model. This silently assumes a symmetry in categories against which both private beliefs can be interpreted: A symmetry which cannot be assumed for human-robot interaction. Hence, on such a model, there is no common ground between humans and robots. This paper presents an alternative model theory, which captures the inherently subjective nature of experience. It is based on a notion of propositions-as-proofs, turning subjective interpretation as well as the projection to intersubjective verification into a notion of inference or argumentation. Beliefs are arguments, whether private, attributed or shared. This results in a formulation of common ground as a dynamic structure of always argued, but possibly only partially confirmed or partially assumed beliefs.

Relation to WP Common ground in dialogue indicates a level of mutual understanding between interlocutors, of what is being talked about. In CogX, "what is being talked about" primarily concerns beliefs about experience of an environment, or about inferred (possibly attributable) characteristics of an environment. Common ground can thus be argued to be based on to the ability to align experience and expectations. This is crucial for an effective transfer of information in communication; without it, dialogue is unlikely to yield insights which the robot can use to drive its learning. The problem that now arises in human-robot interaction is that robots and humans experience reality fundamentally differently. A robot cannot simply assume that a human "symmetrically" understands what it is talking about. The robot needs to reason, within its limited capabilities, how the human may understand the environment, and to what extent that might correspond to how it understands the environment itself. The manuscript describes an approach to formulating an algebraic model theory on which we can define a logic for reasoning about different epistemic and situated perspectives, and how they could be aligned.

2.3 Schoch & Zender, "Situated Plan and Execution Verbalisation"

Bibliography Gerald Schoch and Hendrik Zender. "Situated Plan and Execution Verbalisation." *Technical Report* (2012).

Abstract In this paper, we present an approach to verbalizing reports of intentions and actions of a planner-enabled agent. We consider the case of an interactive intelligent robot that is endowed with a symbolic AI planner. The robot uses the planner to determine and execute sequences of actions in order to achieve a given goal. Since the robot is operating in a real, physical environment, making use of possibly imperfect sensing and actuating, it is likely to encounter unforeseen events or failures, and therefore needs to re-plan in order to come up with alternative plans for achieving its goal. In order to tell a human about what was planned, what was and was not successfully done, what happened – and why – we present a method for generating natural language reports based on such intended plans and the event structure of their execution. This verbalization is inherently situated in nature. For one, external entities that are used in the planning process refer to things, persons, or locations in the physical environment that the robot and the hearer are situated in. Secondly, the robot's intentions and actions are temporally related to the discourse situation: the robot might report about its current plans and attempts for executing an ongoing task; the robot might explain what it did and what it couldn't successfully do in order to establish common ground for asking the human for help; or the the robot might simply report about past events and give details about why it chose to do what.

Relation to WP The approach is integrated in the general natural language processing system developed in this WP and shares its linguistic resources with the other dialogue capabilities, cf. Section 2.4. The verbalization module is connected to the spatial representations developed in WP3 in order to refer to entities in Dora's spatial environment. By this it builds upon previous work in this WP on spatially situated generation of referring expressions.

The approach has been implemented on the integrated robotic system Dora (WP7), which is equipped with a switching symbolic-probabilistic AI planner, able of continual planning and re-planning. Thereby, the work presented here has also a close connection to the research on planning performed in WP4.

2.4 Janíček, "Robust situated language processing with TAROT: The Talking Robots Toolkit"

Bibliography Miroslav Janíček. "Robust situated language processing with TAROT: The Talking Robots Toolkit." Manuscript. (2012).

Abstract This document describes TAROT the Talking Robots Toolkit. TAROT is an open-source software framework for building spoken dialogue functionality for human-robot interaction. TAROT does not impose a specific (cognitive) architecture for building a dialogue system. Its framework allows for multi-threaded (or asynchronous) processing. Processes are defined as open "glass" boxes, and can interact in an event-driven fashion. TAROT is written in the Scala programming language and targets the Java platform.

Relation to WP The manuscript describes a toolkit which consolidates the knowledge and experience gained in CogX in developing systems for situated dialogue processing in HRI.

2.5 Hanheide et al., "Expectations, Intentions, and Actions in Human-Robot Interaction"

Bibliography Marc Hanheide, Manja Lohse and Hendrik Zender. "Expectations, Intentions, and Actions in Human-Robot Interaction." *International Journal of Social Robotics*, 4(2):107–108, Springer Verlag, April 2012.

Abstract Human-robot interaction is becoming increasingly complex through the growing number of abilities, both cognitive and physical, available to today's robots. At the same time, interaction is still often difficult because the users do not understand the robots' internal states, expectations, intentions, and actions. Vice versa, robots lack understanding of the users' expectations, intentions, actions, and social signals.

This article constitutes the editorial of a special issue on "Expectations, Intentions & Actions" of the International Journal of Social Robotics. The special issue bundles recent advances in addressing these challenges. The key questions are how meeting or failing the user's expectations influences the efficiency and effectiveness of human-robot interaction; how more effective and efficient interaction with humans can be achieved using modalities available to a robot; how robots can be equipped with models enabling them to understand their users' state of mind; and similarly, how they can make their own expectations and states explicit through eligible communication channels.

Each of the seven articles in the special issue highlights different aspects around the central theme of *expectations, intentions, and actions in human-robot interaction*. The topics covered range from recognition of verbal and non-verbal cues of intentions and expectations, to verbalization and presentation techniques that make internal processing of the robot accessible to the human.

Relation to WP Meeting or failing to meet the users expectations influences the efficiency and effectiveness of human-robot interaction. The article's discussion complements the more fundamental issues of common ground in HRI presented in Annex 2.2.

References

[1] J.F. Allen and D.J. Litman. Discourse processing and common sense plans. In P.R. Cohen, J. Morgan, and M. Pollack, editors, *Intentions and Communication*. The MIT Press, 1990.

[2] J.F. Allen and C.R. Perrault. Analyzing intention in utterances. *Artificial Intelligence*, 15(3), 1980.

[3] N. Blaylock and J. Allen. A collaborative problem-solving model of dialogue. In L. Dybkjr and W. Minker, editors, *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, pages 200–211, 2005.

[4] M. Brenner. Situation-aware interpretation, planning and execution of user commands by autonomous robots. In *Robot and Human interactive Communication*, The 16th IEEE International Symposium, pages 540–545, 2007.

[5] M. Brenner. Creating dynamic story plots with continual multiagent planning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, july 2010.

[6] M. Brenner and B. Nebel. Continual planning and acting in dynamic multiagent environments. *Journal of Autonomous Agents and Multiagent Systems*, 19(3):297–331, 2009.

[7] H. H. Clark and S. A. Brennan. Grounding in communication. In L.B. Resnick, J.M. Levine, and S.D. Teasley, editors, *Perspectives on socially shared cognition*, pages 127–149. APA Books, Washington VA, 1991.

[8] H. H. Clark and C. R. Marshall. Reference diaries. In D. L. Waltz, editor, *Theoretical issues in natural language processing*, volume 2, pages 57–63. Association for Computing Machinery, New York NY, 1978.

[9] G. Ferguson and J. Allen. Mixed-initiative dialogue systems for collaborative problem-solving. *AI Magazine*, 28(2):23–32, 2007.

[10] R. E. Fikes and N. J. Nilsson. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2(3–4):189–208, 1971.

[11] J. Ginzburg. Situation semantics and the ontology of natural language. In P. Portner, C. Maierborn, and K. von Heusinger, editors, *The Handbook of Semantics*, pages 830–851. de Gruyter, 2011.

[12] M. Göbelbecker, C. Gretton, and R. Dearden. A switching planner for combined task and observation planning. In *Electronic Proceedings of the Workshop on Decision Making in Partially Observable, Uncertain*

Worlds: Exploring Insights from Multiple Communities at the Twenty-Second International Join Conference on Artificial Intelligence (DM-POUW 2011), 2011.

[13] M. Göbelbecker, T. Keller, P. Eyerich, M. Brenner, and B. Nebel. Coming up with good excuses: What to do when no plan can be found. In *Proceedings of the 20th International Conference on Automated Planning and Scheduling (ICAPS)*. AAAI Press, may 2010.

[14] B. J. Grosz and C. L. Sidner. Attention, intention and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[15] B.J. Grosz and L. Hunsberger. The dynamics of intention in collaborative activity. *Cognitive Systems Research*, 7(2-3):259–272, 2006.

[16] B.J. Grosz and S. Kraus. The evolution of sharedplans. In A. Rao and M. Wooldridge, editors, *Foundations and Theories of Rational Agency*, pages 227–262. Springer, 1999.

[17] M. Hanheide, C. Gretton, R. Dearden, N. Hawes, J. Wyatt, A. Pronobis, A. Aydemir, M. Göbelbecker, and H. Zender. Exploiting probabilistic knowledge under uncertain sensing for efficient robot behaviour. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI-11)*, Barcelona, Catalonia, Spain, July 2011.

[18] X. Huang and A. Fiedler. Proof verbalization as an application of nlg. In *Proceedings oft the 15TH International Joint Conference on Artificial Intelligence (IJCAI*, pages 965–970. Morgan Kaufmann, 1997.

[19] M. Janíček. Tarot: The Talking Robots Toolkit: technical description, version 0.1. Technical report, German Research Center for Artificial Intelligence (DFKI), 2012. CogX project deliverable DR6.5.

[20] S. Kiesler. Fostering common ground in human-robot interaction. In *Proceedings of 14th IEEE International Workshop on Robot and Human Interactive Communication (ROMAN 2005)*, 2005.

[21] G. Klein, P.J. Feltovich, J.M. Bradshaw, and D.D. Woods. Common ground and coordination in joint activity. In W.B. Rouse and K.R. Boff, editors, *Organizational Simulation*, pages 139–184. John Wiley, New York City, NY, 2004.

[22] G. Klein, D.D. Woods, J.M. Bradshaw, R. Hoffman, and P. Feltovich. Ten challenges for making automation a team player in joint human-agent activity. *IEEE Intelligent Systems*, 19(6):91–95, November-December 2004.

[23] G.J.M. Kruijff, M. Janíček, and P. Lison. Continual processing of situated dialogue in human-robot collaborative activities. In *Proceedings of the 19th International Symposium on Robot and Human Interactive Communication (RO-MAN 2010)*. IEEE, 2010.

[24] G.J.M. Kruijff, M. Janíček, and H. Zender. Situated communication for joint activity in human-robot teams. *IEEE Intelligent Systems*, March/April 2012. forthcoming.

[25] 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, and J. L. Wyatt, editors, *Cognitive Systems*, volume 8 of *Cognitive Systems Monographs, COSMOS*, chapter 8, pages 311–364. Springer Verlag, Berlin/Heidelberg, Germany, 2010.

[26] B. Loenneker. Narratological knowledge for natural language generation, 2005.

[27] R.R. Murphy and J.L. Burke. The safe human-robot ratio. In M.J. Barnes and F. Jentsch, editors, *Human-Robot Interactions in Future Military Operations*, Human Factors in Defence, pages 31–49. Ashgate, 2010.

[28] M.J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–225, 2004.

[29] M. Poesio and D. Traum. Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309–347, 1997.

[30] C. L. Sidner, C. Lee, C. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.

[31] D. Skočaj, M. Kristan, A. Vrečko, M. Mahnič, M. Janíček, G.J.M. Kruijff, M. Hanheide, N. Hawes, T. Keller, M. Zillich, and K. Zhou. An integrated system for interactive learning in dialogue with a tutor. *under submission*, 2012.

[32] D. Sofge, M. Bugajska, J. G. Trafton, D. Perzanowski, S. Thomas, M. Skubic, S. Blisard, N.L. Cassimatis, D. Brock, W. Adams, and A. Schultz. Collaborating with humanoid robots in space. *International Journal of Humanoid Robotics*, 2(2):181–201, 2005.

[33] M. Stone. Intention, interpretation and the computational structure of language. *Cognitive Science*, 28(5):781–809, 2004.

[34] M. Stone and R.H. Thomason. Context in abductive interpretation. In *Proceedings of EDILOG 2002: 6th workshop on the semantics and pragmatics of dialogue*, 2002.

[35] 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.

[36] K. Stubbs, P.J. Hinds, and D. Wettergreen. Autonomy and common ground in human-robot interaction: A field study. *IEEE Intelligent Systems*, 22:42–50, March 2007. Special Issue on Interacting with Autonomy.

[37] I. M. T. Swartjes and M. Theune. The virtual storyteller: story generation by simulation. In *Proceedings of the Twentieth Belgian-Netherlands Conference on Artificial Intelligence, BNAIC 2008, Enschede*, pages 257–265, Enschede, October 2008. University of Twente.

[38] J. G. Trafton, N. L. Cassimatis, M. D. Bugajska, D. P. Brock, F. E. Mintz, and A. C. Schultz. Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 35(4):460–470, 2005.

[39] D. Wilkes-Gibbs and H. H. Clark. Coordinating beliefs in conversation. *Journal of Memory and Cognition*, 31:183–194, 1992.

[40] H. Zender. *Situated Production and Understanding of Verbal References to Entities in Large-Scale Space*, volume 36 of *Saarbrücken Dissertations in Computational Linguistics and Language Technology.* German Research Center for Artificial Intelligence and Saarland University, Saarbrücken, Germany, 2011.

[41] H. Zender. Multi-layered conceptual spatial mapping – representing spatial knowledge for situated action and human-robot interaction. In Y. Amirat, A. Chibani, and G. P. Zarri, editors, *Bridges Between the Methodological and Practical Work of the Robotics and Cognitive Systems Communities – From Sensors to Concepts*, Intelligent Systems Reference Library. Springer Verlag, Berlin/Heidelberg, Germany, to appear 2012.

[42] H. Zender, G.J.M. Kruijff, and I. Kruijff-Korbayová. Situated resolution and generation of spatial referring expressions for robotic assistants. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, pages 1604–1609, Pasadena, CA, USA, July 2009.

Abductive Reasoning for Continual Dialogue Understanding

Miroslav Janíček

German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
miroslav.janicek@dfki.de

Abstract. In this paper we present a continual context-sensitive abductive framework for understanding situated spoken natural dialogue. The framework builds up and refines a set of partial defeasible explanations of the spoken input, trying to infer the speaker's intention. These partial explanations are conditioned on the eventual verification of the knowledge gaps they contain. This verification is done by executing test actions, thereby going beyond the initial context. The approach is illustrated by an example set in the context of human-robot interaction.

Keywords: Intention recognition, natural language understanding, abduction, context-sensitivity.

1 Introduction

In task-oriented dialogues between two agents, such as between two humans or a human and a robot, there is more to dialogue than just understanding words. The robot needs to understand what is being talked about, but it also needs to understand why it was told something. In other words, what the human *intends* the robot to do with the information in the larger context of their joint activity.

Therefore, understanding language can be phrased as an *intention recognition* problem: given an utterance from the human, how do we find the intention behind it?

In this paper, we explore an idea inspired by the field of continual planning [8], by explicitly capturing the possible knowledge gaps in such an interpretation. The idea is based on the notion of *assertion*, an explicit test for the validity of a certain fact, going beyond the current context.

The structure of the paper is as follows. After briefly introducing the notion of intention recognition, abduction and situatedness in the next section, we introduce the continual abductive reasoning mechanism in §3, and discuss it on an example in §4, before concluding with a short summary.

2 Background

The idea of expressing *understanding* in terms of intention recognition has been introduced by H. P. Grice [12,20]. In this paper, we build on Stone and Thomason's approach to the problem [23] who in turn extend the work done by Hobbs and others [13], and base their approach to intention recognition on *abductive reasoning*.

2.1 Abduction

Abduction is a method of explanatory logical reasoning introduced into modern logic by Charles Sanders Peirce [11]. Given a theory T, a rule $T \vdash A \rightarrow B$ and a fact B, abduction allows inferring A as an explanation of B. B can be deductively inferred from $A \cup T$. If $T \nvdash A$, then we say that A is an *assumption*.

There may be many possible causes of B besides A. Abduction amounts to *guessing*; assuming that the premise is true, the conclusion holds too. To give a well-known example:

> Suppose we are given two rules saying "if the sprinkler is on, then the lawn is wet" and "if it rained, then the lawn is wet". Abductively inferring the causes for the fact that the lawn is wet then yields two possible explanations: the sprinkler is on, or it rained.

Obviously, as there may be many possible explanations for a fact, in practical applications there needs to be a mechanism for selecting the best one. This may be done by purely syntactic means (e.g. lengths of proofs), or semantically by assigning *weights* to abductive proofs and selecting either the least or most costly proof [22], or by assigning probabilities to proofs [18]. In that case, the most probable proof is also assumed to be the best explanation. Our approach combines both aspects.

2.2 Intention Recognition

Abduction is a suitable mechanism to perform inferences on the pragmatic (discourse) level. For understanding, abduction can be used to infer the explanation *why* an agent said something, in other words the *intention* behind the utterance.

An intention is usually modelled as a goal-oriented cognitive state distinct from *desires* in that there is an explicit commitment to acting towards the goal and refraining from actions that may render it impossible to achieve [7,10].

For the purposes of this paper, we shall treat intentions as intended actions that have pre- and post-conditions, similar to planning operators in automated planning. Pre-conditions express the necessary conditions *before* the action is executed (and sufficient for its execution), and post-conditions express the necessary conditions *after* the action is executed.

Note that reasoning with intentions allows us to reverse the task, and search for appropriate (surface) presentation of a given intention [24]. Intentions can therefore serve as a middle representational layer and abduction as the inference mechanism by using which we either turn a realisation into an intention, or the other way around.

2.3 Situated Understanding

Suppose that a human user is dealing with a household robot capable of manipulating objects (finding them, picking them up, putting them down). The human wants the robot to bring him the mug from the kitchen, so he instructs the robot by saying:

"Bring me the mug from the kitchen."

Now, what should the robot do? In the beginning, the utterance is just a stream of audio. The robot has to detect voice in the audio data, and if the speech recognition works well enough, it will be able to obtain the surface form of the utterance, i.e. the words that were spoken by the human.

Once the word sequence is recognised, the robot needs to assign linguistic structure to it so that it can reason about its logical structure. The logical structure of the utterance is typically not in any way related to the actual *situated* experience of the robot. The noun phrases "the mug" and "the kitchen" are just referring expressions *standing for* some entities in the real world, and can be manipulated as expressions using logical rules without the need to be concerned about value of the standing-for relation.

However, this relation is absolutely crucial to understanding what the human said and why. Without being able to reduce the referring expressions to the corresponding real-world entities there is no true understanding, and – more importantly – there can be no appropriate reaction, which presumably is one of the reasons why the human uttered the sentence in the first place (i.e. to elicit such a reaction).

Grounding the relation in reality is therefore a crucial task that any cognitive agent has to tackle. However, since all sensory perception is necessarily partial and subject to uncertainty, there is no guarantee that the "knowledge base", a formalisation of the current snapshot of the knowledge about the world, contains the information necessary for such a grounding. In other words, a situated agent cannot afford the luxury of reasoning under closed-world assumption, and has to venture beyond that.

This means that the robot must be able identify its knowledge gaps, and verify or falsify them *while* trying to understand the human's utterance. This implies that the processes of understanding an input and acting on it are interleaved and that there is a bi-directional interface between them.

3 Approach

This paper extends the work of Stone and Thomason on context-sensitive language understanding by explicitly modelling the knowledge gaps that inevitably arise in such an effort due to uncertainty and partial observability. The approach is based on generating partial hypotheses for the explanation of the observed behaviour of other agents, under the assumption that the observed behaviour is intentional. These partial hypotheses are defeasible and conditioned on the validity (and eventual verification) of their assumptions.

In this section, we examine the abductive reasoning system capable of representing knowledge gaps in the form of partial proofs, how such partial proofs can be generated and verified or falsified, and the semantic framework used in our system to capture linguistic meaning that the system then grounds in reality.

3.1 Partial Abductive Proofs

Our abductive inference mechanism is essentially Hobbs and Stickel's logic programming approach to weighted abduction [13,22] enhanced by a contextual aspect [3] with the weights in the system being assigned a probabilistic interpretation following Charniak and Shimony [9].

Abduction Context. Inference in our system makes use of four ingredients: *facts* (denoted $\mathcal{F}$), *rules* ($\mathcal{R}$), *disjoint declarations* ($\mathcal{D}$) and *assumability functions* ($\mathcal{S}$), collectively called the *abduction context.* The proof procedure uses these iteratively in order to derive proofs of an initial *goal.*

- *Facts* are modalised formulas of the form

$$\mu : A$$

  where μ is a (possibly empty) sequence of modal contexts, and A is an atomic formula, possibly containing variables.
- *Rules* are modalised Horn clauses, i.e. formulas of the form

$$(\mu_1 : A_1/t_1) \wedge ... \wedge (\mu_n : A_n/t_n) \rightarrow (\mu_H : H)$$

  where each of the $\mu_i : A_i$ and $\mu_H : H$ are modalised formulas. Each antecedent is annotated by t_i, which determines the way the antecedent is manipulated and is one of the following:
  - *assumable(f)* – the antecedent is assumable under function f;
  - *assertion* – the antecedent is asserted, i.e. identifies a knowledge gap, conditioning the validity of the proof on it being proved in a subsequent reinterpretation (see below).
- *Assumability functions* are partial functions f, $f : \mathcal{P}(\mathcal{F}) \rightarrow \mathbb{R}_0^+$, where $\mathcal{P}(\mathcal{F})$ is the set of modalised formulas, with the additional monotonicity property that if $F \in \mathrm{dom}(f)$, then for all more specific (in terms of variable substitution) facts F', $F' \in \mathrm{dom}(f)$ and $f(F) \leq f(F')$. We also define an empty ("truth") assumability function $\perp$ such that $\mathrm{dom}(\perp) = \emptyset$.

  Since they are partial functions, assumability functions determine both whether a modalised formula may be assumed and the cost of such an assumption. As a special case, the empty assumability function $\perp$ can be used to prevent the formula from being assumed altogether.

- A *disjoint declaration* is a statement of the form

$$\text{disjoint}(\mu : A_1, \ldots, \mu : A_n)$$

  which specifies that at most one of the modalised formulas $\mu : A_i$ may be used in the proof. A_i and A_j cannot be unified for all $i \neq j$.

Proof Procedure. The proof procedure is an iterative rewriting process starting from some initial *goal* state. A *proof state* is a sequence of marked modalised formulas (called *queries* in this context)

$$Q_1[n_1], \ldots, Q_m[n_m]$$

The markings n_i are one of the following:

- *unsolved(f)* – the query is yet to be proved and can be assumed if it is in the domain of the assumability function f;
- *proved* – the query is proved in the proof state;
- *assumed(f)* – the query is assumed under assumability function f;
- *asserted* – the query is asserted – its validity is not to be determined in the current context.

Algorithm 1 defines the proof procedure in detail. The top-level function ABDUCE takes an abduction context c and a proof state Π, and returns a set of proof states that

(1) are transformations of Π,
(2) are consistent with c, and
(3) do not contain any *unsolved* queries.

First, the input proof state is checked for validity with respect to the disjoint declarations $\mathcal{D}$ in the function IS-DISJOINT-VALID. If the check turns out to be negative, the proof state is discarded, and ABDUCE returns an empty set.

If Π satisfies the disjointness constraints, the function TF-DUP turns it into a set of proof states where unsolved queries that have already been proved, assumed or asserted are removed. The transformation returns a non-empty set of proof states. This step ensures that no query is examined more than once.

Next, each proof state resulting from TF-DUP is again checked whether it contains an unsolved query. If it does not, then the conditions (1)–(3) above are already fulfilled, and the proof state ends up in the result.

If it does, the proof procedure resolves the proof state against the facts, rules and assumability functions, collecting the results, and recursively calling ABDUCE on them so as to satisfy the above conditions.

Formally, given a proof state

$$\Pi = Q_1[n_1], \ldots, Q_m[n_m]$$

where Q_i is the leftmost query marked (guaranteed to exist at this point) as *unsolved*(f) where f is an assumability function, the transformation rules TF-FACT, TF-RULE and TF-ASSUME each return a (possibly empty) set of transformed proof states, and are defined as follows:

– TF-FACT (resolution with a fact): For all $Q \in \mathcal{F}$ such that the Q and Q_i are unifiable with a most general unifier σ (denoted $\sigma = \mathrm{unify}(Q, Q_i)$), add a new state Π' to the result of the transformation:

$$\Pi' = Q_1\sigma[n_1], \dots, Q_i\sigma[\mathit{proved}], \dots Q_m\sigma[n_m]$$

– TF-RULE (resolution with a rule): For each rule $r \in \mathcal{R}$ of the form

$$G_1/t_1, \dots, G_k/t_k \to H$$

(with variables renamed so that it has no variables in common with Π) such that there is a $\sigma = \mathrm{unify}(H, Q_i)$, i.e. the rule head is unifiable with the unsolved query, add a new state Π' to the transformation result:

$$\begin{aligned}\Pi' = {} & Q_1\sigma[n_1], \dots, Q_{i-1}\sigma[n_{i-1}],\\ & G_1\sigma[p_1], \dots, G_k\sigma[p_k], Q_i\sigma[\mathit{proved}],\\ & Q_{i+1}\sigma[n_{i+1}], \dots, Q_m\sigma[n_m]\end{aligned}$$

The query markings p_i are derived from t_i for all $i \in \{1, \dots, k\}$ as follows:

$$\begin{aligned}&\text{if } t_i = \mathit{assumable}(f), \text{ then } p_i = \mathit{unsolved}(f)\\ &\text{if } t_i = \mathit{assertion}, \qquad \text{then } p_i = \mathit{asserted}\end{aligned}$$

– TF-ASSUME (assumption): If $Q \in \mathrm{dom}(f)$ such that there is a most general unifier $\sigma = \mathrm{unify}(Q, Q_i)$, add a new state Π' to result of the transformation:

$$\Pi' = Q_1\sigma[n_1], \dots, Q_i\sigma[\mathit{assumed}(f)], \dots, Q_n\sigma[n_m]$$

Note that the proof procedure along with the definition of assumability functions ensures that the cost of the proofs are monotonic with respect to unification and rule application, allowing for the use of efficient search strategies.

Knowledge Gaps and Assertions. Our extension of the "classical" logic-programming-based weighted abduction as proposed by Stickel and Hobbs lies in the extension of the proof procedure with the notion of *assertion* based on the work in continual automated planning [8], allowing the system to reason about information not present in the knowledge base, thereby addressing the need for reasoning under the open-world assumption.

In continual automated planning, assertions allow a planner to reason about information that is not known at the time of planning (for instance, planning for information gathering), an assertion is a construct specifying a "promise" that the information in question will be resolved eventually. Such a statement requires planning to be a step in a continual loop of interleaved planning and acting.

By using a logic programming approach, we can use unbound variables in the asserted facts in order to reason not only about the fact that the given assertion will be a fact, but also under-specify its eventual arguments.

Algorithm 1 Weighted abduction

```
ABDUCE(c = (𝓕, 𝓡, 𝓓, 𝓢), Π = Q_1[n_1], ..., Q_m[n_m]):
  if IS-DISJOINT-VALID(𝓓, Π) then
    R ← ∅
    for all Π′ ∈ TF-DUP(Π) do
      if Π′ contains a query marked as unsolved then
        H ← TF-FACT(𝓕, Π′) ∪ TF-RULE(𝓡, Π′) ∪ TF-ASSUME(𝓢, Π′)
        R ← R ∪ ⋃_{P∈H} ABDUCE(c, P)
      else
        R ← R ∪ {Π′}
      end if
    end for
    return R
  else
    return ∅
  end if

IS-DISJOINT-VALID(𝓓, Π = Q_1[n_1], ..., Q_m[n_m]):
  for all d = disjoint(D_1, ..., D_q) ∈ 𝓓 do
    if ∃i ≠ j ≠ k ≠ l s.t. ∃σ, σ′: σ = unify(D_i, Q_k) and σ′ = unify(D_j, Q_l) then
      return false
    end if
  end for
  return true

TF-DUP(Π = Q_1[n_1], ..., Q_m[n_m]):
  if Π contains a query marked as unsolved then
    i ← arg min_{j∈{1,...,m−1}}(∃f s.t. n_j = unsolved(f))
    H ← ∅
    for all s ∈ {i + 1, ..., m} s.t. unify(Q_i, Q_s) = σ do
      H ← H ∪ TF-DUP(Q_1σ[n_1], ..., Q_{i−1}σ[n_{i−1}], Q_{i+1}σ[n_{i+1}] ..., Q_mσ[n_m])
    end for
    if H ≠ ∅ then return H else return {Π} end if
  else
    return {Π}
  end if
```

The proposed notion of *assertion* for our abductive system is based on *test actions* $\langle F \rangle$ [4]. Baldoni et al. specify a test as a proof rule. In this rule, a goal F follows from a state $a_1, ..., a_n$ after steps $\langle F \rangle, p_1, ..., p_m$ if we can establish F on $a_1, ..., a_n$ with answer σ and this (also) holds in the final state resulting from executing $p_1, ..., p_m$.

An assertion is the transformation of a test into a partial proof which assumes the verification of the test, while at the same time conditioning the obtainability of the proof goal on the tested statements. $\mu : \langle D \rangle$ within a proof $\Pi[\langle D \rangle]$ to a goal C turns into $\Pi[D] \to C \wedge \mu : D$. Should $\mu : D$ not be verifiable, Π is invalidated.

Probabilistic Interpretation. In weighted abduction, weights assigned to assumed queries are used to calculate the overall proof cost. The proof with the lowest cost is the best explanation. However, weights are usually not assigned any semantics, and often a significant effort by the writer of the rule set is required to achieve expected results [13].

However, Charniak and Shimony [9] showed that by setting weights to $-\log$ of the prior probability of the query, the resulting proofs can be given probabilistic semantics.

Suppose that query Q_k can be assumed true with some probability $P(Q_k \text{ is true})$. Then if Q_k is assumable under assumability function f such that $f(Q_k) = -\log(P(Q_k \text{ is true}))$, and under the independence assumption, we can represent the overall probability of the proof $\Pi = Q_1[n_1], ..., Q_n[n_m]$ as

$$P(\Pi) = e^{\sum_{k=1}^{m} cost(Q_k)}$$

where

$$cost(Q_k) = \begin{cases} f(Q_k) & \text{if } n_k = assumed(f) \\ 0 & \text{otherwise} \end{cases}$$

The best explanation Π_{best} of a the goal state G is then

$$\Pi_{best} = \operatorname*{arg\,min}_{\Pi \text{ proof of } G} P(\Pi)$$

Exact inference in such a system is NP-complete, and so is approximate inference given a threshold [9]. However, it is straightforward to give an anytime version of the algorithm – simply by performing iterative deepening depth-first search [19] and memorising a list of most probable proofs found so far.

Comparison with Other Approaches. Our system is similar to Poole's Probabilistic Horn abduction [18]. The main difference, apart from the proof procedure which is cost-based in our case, is that we do not include probabilities in our formulation of disjoint declarations. Since we avoid duplicate assumptions, we are able to model the semantics of disjoint declarations with probabilities.

On the other hand, having a general disjoint declaration allows us to define exclusivity rules such as

Algorithm 2 (Nondeterministic) continual abduction

```
CONTINUAL-ABDUCTION(c, Π):
  c = context
  Π = proof

  while Π contains assertion A do
    c′ ← TEST-ACTION(c, A)
    H ← ABDUCE(c′, A)
    for all Π′ ∈ H do
      CONTINUAL-ABDUCTION(c′, Π′)
    end for
  end while
  return Π
```

$$\text{disjoint}([\text{p}(X, \text{yes}), \text{p}(X, \text{no})])$$

without having to specify the prior probabilities of the disjuncts.

Moreover, in our rule sets for natural language understanding and generation, we need to be able to manipulate logical structure (e.g. logical forms of utterances) efficiently. We have found that the logic-programming-based approach is quite satisfactory in this aspect, since it permits the use of standard Prolog programming techniques. In approaches to probabilistic abduction that are *not* based on logic programming, such as Kate and Mooney's abduction in Markov Logic Networks [15], these techniques are not applicable, which crucially limits their applicability to our domain.

3.2 Generating Partial Hypotheses

For each goal G, a the function ABDUCE returns a set of proofs H, with a total ordering on this set. Due to the use of assertions, some of these proofs may be partial, and their validity has to be verified. The presence of assertions in the proofs means that there is a knowledge gap, namely the truth value of the assertion. Each assertion thus specifies the need for performing a (test) action. This action might require the access to other knowledge bases than the abductive context, as in the case of resolving referring expressions, or an execution of a physical action.

Formally, given an initial goal G and context c, the abduction procedure produces a set H of hypotheses $c : \Pi \rightarrow C \wedge c_i : A_i$, where c_i is a sub-context in which where an assertion $A_i \in \Pi$ may be evaluated. Such proofs are thus both *partial* and *defeasible* — they may be both extended and discarded, depending on the evaluation of the assertions.

The set of possible hypotheses is continuously expanded until the best full proof is found. This process is defined in Algorithm 2.

The algorithm defines the search space in which it is possible to find the most probable proof of the initial goal G. The important point is, however, that it is just that — a definition. The actual implementation may keep track of the partial hypotheses it defines, and take the appropriate test actions when necessary, or postpone them indefinitely. The cost of *performing* an action is not factored into the overall proof cost.

The partial hypotheses therefore serve as an interface layer between the language understanding and external decision-making processes (such as planning in a robotic architecture).

3.3 Representing Linguistic Meaning

For representing linguistic meaning in our system we use the *Hybrid Logic Dependency Semantics* (HLDS), a hybrid logic framework that provides the means for encoding a wide range of semantic information, including dependency relations between heads and dependents [21], tense and aspect [17], spatio-temporal structure, contextual reference, and information structure [16].

Hybrid Logic. Classical modal logic suffers from a surprising "asymmetry". Although the concept of states ("worlds") is at the heart of model theory, there is no way to directly reference specific states in the object language. This asymmetry is at the root of several theoretical and practical problems facing modal logic [6,1].

Hybrid logic provides an elegant solution to many of these problems. It extends standard modal logic with *nominals*, another sort of basic formulas that explicitly name worlds in the object language. Next to propositions, nominals—and, by extension, possible worlds—therefore become first-class citizens in the object language. The resulting logical framework retains decidability and favourable complexity [2].

Each nominal names a unique state. To get to that state, a new operator is added, the *satisfaction operator*. The satisfaction operator that enables us to "jump" to the state named by a nominal. The satisfaction operator is written $@_i$, where i is a nominal.

Formally, let $\textsc{Prop} = \{p, q, ...\}$ be a set of propositional symbols, $\textsc{Mod} = \{\pi, \pi', ...\}$ a set of modality labels, and $\textsc{Nom} = \{i, j, ...\}$ a non-empty set disjoint from $\textsc{Prop}$ and $\textsc{Mod}$. We define the well-formed formulas of the basic hybrid multimodal language $\mathcal{L}_@$ over $\textsc{Prop}$, $\textsc{Mod}$ and $\textsc{Nom}$ as such:

$$\phi ::= i \mid p \mid \neg\phi \mid \phi \rightarrow \varphi \mid \langle\pi\rangle\phi \mid [\pi]\phi \mid @_i\phi$$

A formula $@_i\phi$ states that the formula ϕ holds at the unique state named by i. In more operational terms, the formula $@_i\phi$ could be translated in the following way: "go to the (unique!) state named by i, and check whether ϕ is true at that state".

Hybrid Logic Dependency Semantics. HLDS uses hybrid logic to capture dependency complexity in a model-theoretic relational structure, using ontological sorting to capture categorial aspects of linguistic meaning, and naturally capture (co-)reference by explicitly using *nominals* in the representation. The dependency structures can be derived from CCG [5], which is the setup used in our system, but other approaches are possible.

Generally speaking, HLDS represents an expression's linguistic meaning as a conjunction of modalised terms, anchored by the nominal that identifies the head's proposition:

$$@_{h:\mathsf{sort}_h}\ (\mathbf{prop}_h \wedge \langle \mathsf{R}_i \rangle\ (d_i : \mathsf{sort}_{d_i} \wedge \mathbf{dep}_i))$$

Here, the head proposition nominal is h. $\mathbf{prop}_h$ represents the *elementary predication* of the nominal h. The dependency relations (such as Agent, Patient, Subject, etc.) are modelled as modal relations $\langle R_i \rangle$, with the dependent being identified by a nominal d_i. Features attached to a nominal (e.g. $\langle \mathsf{Num} \rangle$ $\langle \mathsf{Quantification} \rangle$, etc.) are specified in the same way.

Figure 1 gives an example of HLDS representation (logical form) of the sentence "Bring me the mug from the kitchen". The logical form has six nominals, $event_1$, $agent_1$, $person_1$, $thing_1$, $from_1$ and $thing_1$, that form a dependency structure: $event_1$ is the the head of dependency relations Actor (the dependent being $agent_1$), Patient ($thing_1$), Recipient ($person_1$), Modifier ($from_1$), and Subject (the sentence subject, $agent_1$).

Each nominal has an ontological sort (illustrated on $event_1$, the sort is action-non-motion) a proposition (**bring**), and may have features (Mood).

$$
\begin{array}{l}
@_{event_1:\mathsf{action\text{-}non\text{-}motion}}(\mathbf{bring} \wedge \\
\quad \langle \mathsf{Mood} \rangle\ \mathbf{imp} \wedge \\
\quad \langle \mathsf{Actor} \rangle\ (agent_1 : \mathsf{entity} \wedge \mathbf{addressee}) \\
\quad \langle \mathsf{Patient} \rangle\ (thing_1 : \mathsf{thing} \wedge \mathbf{mug} \wedge \\
\qquad \langle \mathsf{Delimitation} \rangle\ \mathbf{unique} \wedge \\
\qquad \langle \mathsf{Num} \rangle\ \mathbf{sg} \wedge \\
\qquad \langle \mathsf{Quantification} \rangle\ \mathbf{specific}) \wedge \\
\quad \langle \mathsf{Recipient} \rangle\ (person_1 : \mathsf{person} \wedge \mathbf{I} \wedge \\
\qquad \langle \mathsf{Num} \rangle\ \mathbf{sg}) \wedge \\
\quad \langle \mathsf{Modifier} \rangle\ (from_1 : \mathsf{m\text{-}wherefrom} \wedge \mathbf{from} \wedge \\
\qquad \langle \mathsf{Anchor} \rangle\ (place_1 : \mathsf{e\text{-}place} \wedge \mathbf{kitchen} \wedge \\
\qquad\quad \langle \mathsf{Delimitation} \rangle\ \mathbf{unique} \wedge \\
\qquad\quad \langle \mathsf{Num} \rangle\ \mathbf{sg} \wedge \\
\qquad\quad \langle \mathsf{Quantification} \rangle\ \mathbf{specific})) \wedge \\
\quad \langle \mathsf{Subject} \rangle\ agent_1 : \mathsf{entity})
\end{array}
$$

Fig. 1. HLDS semantics for the utterance "Bring me the mug from the kitchen"

Every logical form in HLDS, being a formula in hybrid logic, can be decomposed into a set of facts in the abductive context corresponding to its minimal Kripke model. The resulting set of abduction facts obtained by decomposing the logical form in Figure 1 is shown by Figure 3.

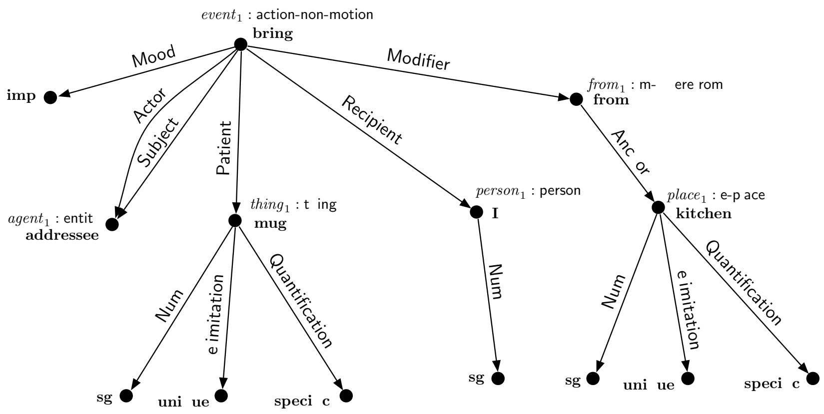

Fig. 2. Minimal model for the hybrid logic formula in Figure 1

HLDS only represents the meaning as derived from the linguistic realisation of the utterance and does not evaluate the state of affairs denoted by it. This sets the framework apart from semantic formalisms such as DRT [14]. The grounding in reality is partly provided by the continual abductive framework by generating and validating (or ruling out) partial abductive hypotheses as more information is added to the system.

4 Example

Let us examine the mechanism in an example introduced in §2.3.

The human's utterance,

"Bring me the mug from the kitchen."

is analysed in terms of HLDS (see Figure 1), and its translation (see Figure 3) is made part of the abduction context c.

The robot tries to make sense of the utterance by proving the goal

$$\mathrm{uttered}(\mathrm{human}, \mathrm{robot}, \mathrm{event}_1)$$

in the abduction context c. Suppose that the best proof state returned by ABDUCE is the following:

$\mathrm{sort}(\mathrm{event}_1, \text{action-non-motion}),$
$\mathrm{prop}(\mathrm{event}_1, \mathrm{bring}),$
$\mathrm{feat}(\mathrm{event}_1, \mathrm{mood}, \mathrm{imp}),$
$\mathrm{rel}(\mathrm{event}_1, \mathrm{actor}, \mathrm{agent}_1),$
$\mathrm{sort}(\mathrm{agent}_1, \mathrm{entity}),$
$\mathrm{prop}(\mathrm{agent}_1, \mathrm{addressee}),$
$\mathrm{rel}(\mathrm{event}_1, \mathrm{patient}, \mathrm{thing}_1),$
$\mathrm{sort}(\mathrm{thing}_1, \mathrm{thing}),$
$\mathrm{prop}(\mathrm{thing}_1, \mathrm{mug}),$
$\mathrm{feat}(\mathrm{thing}_1, \mathrm{delimitation}, \mathrm{unique}),$
$\mathrm{feat}(\mathrm{thing}_1, \mathrm{num}, \mathrm{sg}),$
$\mathrm{feat}(\mathrm{thing}_1, \mathrm{quantification}, \mathrm{specific}),$
$\mathrm{rel}(\mathrm{event}_1, \mathrm{recipient}, \mathrm{person}_1),$
$\mathrm{sort}(\mathrm{person}_1, \mathrm{person}),$
$\mathrm{prop}(\mathrm{person}_1, \mathrm{i}),$
$\mathrm{feat}(\mathrm{person}_1, \mathrm{num}, \mathrm{sg}),$
$\mathrm{rel}(\mathrm{event}_1, \mathrm{modifier}, \mathrm{from}_1),$
$\mathrm{sort}(\mathrm{from}_1, \text{m-wherefrom}),$
$\mathrm{prop}(\mathrm{from}_1, \mathrm{from}),$
$\mathrm{rel}(\mathrm{from}_1, \mathrm{anchor}, \mathrm{place}_1),$
$\mathrm{sort}(\mathrm{place}_1, \text{e-place}),$
$\mathrm{prop}(\mathrm{place}_1, \mathrm{kitchen}),$
$\mathrm{feat}(\mathrm{place}_1, \mathrm{delimitation}, \mathrm{unique}),$
$\mathrm{feat}(\mathrm{place}_1, \mathrm{num}, \mathrm{sg}),$
$\mathrm{feat}(\mathrm{place}_1, \mathrm{quantification}, \mathrm{specific}),$
$\mathrm{rel}(\mathrm{event}_1, \mathrm{subject}, \mathrm{agent}_1)$

Fig. 3. The translation of the hybrid logic formula in Figure 1 into abduction facts

$$\begin{array}{rr}
\mathrm{uttered}(\mathrm{human}, \mathrm{robot}, \mathrm{event}_1)\ [\mathit{proved}] & (1) \\
\hline
\mathrm{prop}(\mathrm{event}_1, \mathrm{bring})\ [\mathit{proved}] & (2) \\
\mathrm{intends}(\mathrm{event}_1, \mathrm{human}, I)\ [\mathit{assumed}(\mathit{engagement})] & (3) \\
\mathrm{rel}(\mathrm{event}_1, \mathrm{patient}, \mathrm{thing}_1)\ [\mathit{proved}] & (4) \\
\text{refers-to}(\mathrm{thing}_1, X)\ [\mathit{asserted}] & (5) \\
\text{refers-to}(\mathrm{place}_1, P)\ [\mathit{asserted}] & (6) \\
\mathrm{pre}(I, \mathrm{object}(X))\ [\mathit{asserted}] & (7) \\
\mathrm{pre}(I, \text{is-in}(X, P))\ [\mathit{asserted}] & (8) \\
\text{refers-to}(\mathrm{person}_1, \mathrm{human})\ [\mathit{proved}] & (9) \\
\mathrm{prop}(\mathrm{person}_1, i)\ [\mathit{proved}] & (10) \\
\mathrm{rel}(\mathrm{event}_1, \mathrm{recipient}, \mathrm{person}_1)\ [\mathit{proved}] & (11) \\
\mathrm{post}(I, \mathrm{has}(\mathrm{human}, X))\ [\mathit{proved}] & (12)
\end{array}$$

The proof is an explanation of the event (1) in terms of a partially specified intention I (3), defined by its pre- and post-conditions. The pre-conditions are the existence of an entity X (7) and that X is located *in* another entity P (8). The post-condition (12) is the resulting state in which the human has X (12).

The proof appeals to the logical form of the utterance in atoms (2), (4), (10), (11). Also, atom (9) is proved from (1) and (10) (whoever uses "I" refers to themselves), and (12) is a consequence of (2), (9) and (11) (bringing x to a person p means ending up in a state in which p has x).

Atom (3) is assumed under the assumability function *engagement*, which is supplied in the abduction context before calling ABDUCE and specifies the robot's subjective probability of being engaged in a conversation with the particular human at the time the utterance was observed.

Note that the proof state contains four atoms marked as assertions: (5), (6), (7) and (8). These are the explicit gaps in the proof that make it a partial interpretation. They are chosen by the domain engineer, and since they need to be verified (or falsified) by an external process, they form the interface to external knowledge bases and decision-making. Since for now those atoms are marked as asserted, there is nothing more that ABDUCE can do.

The initiative then shifts to an external decision-making process. It selects some of the assertions, and tries to verify them.

A sensible strategy[1] might be to first establish the referent of $place_1$. This could be resolved against the internal knowledge base (in case the robot has been given a tour of the household), or it could trigger the exploration behaviour – in order to resolve $place_1$, the robot could try finding it first. Again, choosing which behaviour is more appropriate depends on the application, and on the planning method that is invoked by the decision-maker in order to verify the assertion.

Once the assertion is verified, the proof is updated accordingly, in our case by replacing all occurrences by replacing the unbound variable P by a unique symbol, for instance by the identifier $id_{kitchen}$ of the topological region in the robot's topological map:

$$\frac{\text{refers-to}(\text{place}_1, id_{kitchen})\ [\textit{proved}] \qquad (6')}{\text{resolves-to-toporegion}(\text{place}_1, id_{kitchen})\ [\textit{assumed}(\textit{topo})]\ (6'')}$$

The atom (6) in the original proof state is expanded by a proof state consisting of queries (6') and (6"), thereby replacing P in the entire proof by $id_{kitchen}$, and adding the cost of assuming (6") to the overall proof cost. This atom is assumed under an assumability function *topo*, supplied as part of the abduction context in which the proof is expanded – i.e. by the external knowledge source. An assumption is added instead of a fact so that the external knowledge base performing this operation can express uncertainty about the resolution result.

The proof is therefore expanded into the following:

[1] Note that the problem of what determing good verification strategies and choosing them is beyond the scope of this paper.

$$\frac{\text{uttered}(\text{human}, \text{robot}, \text{event}_1)\ [\textit{proved}] \quad (1)}{\begin{array}{rlr} \text{prop}(\text{event}_1, \text{bring}) & [\textit{proved}] & (2) \\ \text{intends}(\text{event}_1, \text{human}, I) & [\textit{assumed}(\textit{engagement})] & (3) \\ \text{rel}(\text{event}_1, \text{patient}, \text{thing}_1) & [\textit{proved}] & (4) \\ \text{refers-to}(\text{thing}_1, X) & [\textit{asserted}] & (5) \\ \text{refers-to}(\text{place}_1, id_{kitchen}) & [\textit{proved}] & (6') \\ \text{resolves-to-toporegion}(\text{place}_1, id_{kitchen}) & [\textit{assumed}(\textit{topo})] & (6'') \\ \text{pre}(I, \text{object}(X)) & [\textit{asserted}] & (7) \\ \text{pre}(I, \text{is-in}(X, id_{kitchen})) & [\textit{asserted}] & (8) \\ \text{refers-to}(\text{person}_1, \text{human}) & [\textit{proved}] & (9) \\ \text{prop}(\text{person}_1, i) & [\textit{proved}] & (10) \\ \text{rel}(\text{event}_1, \text{recipient}, \text{person}_1) & [\textit{proved}] & (11) \\ \text{post}(I, \text{has}(\text{human}, X)) & [\textit{proved}] & (12) \end{array}}$$

Now there are just three assertions left: (5), (7) and (8). These express the knowledge gaps about the referent of "the mug", the existence of the object, and its location, respectively.

There are, as before, several possible ways of verifying these. The most sensible one would probably be going to the kitchen (i.e. the topological region $id_{kitchen}$) and searching for objects there, which would verify both (8) and (7) and expand them with all objects it finds. There would be many parallel proof states resulting from such an expansion, and the robot would have to prune them down by verifying the remaining assertion (5).

One way of doing that would be to bring all objects one by one to the human, asking "did you mean this one?" Alternatively, the robot might simply bring the most likely object. The human's acceptance of the choice would then verify the assertion. This is, again, a matter for consideration in the higher level of planning and goal management.

5 Conclusion

This paper presents an abductive framework for natural language understanding that is based on abductive reasoning over partial hypotheses. The framework is set within the process of intention recognition.

The abductive framework is contextually-enhanced version of a logic programming approach to weighted abduction with a probabilistic semantics assigned to the weights. Our extension of this framework is in the introduction of the notion of *assertion*, which is essentially a requirement for a future test to verify or falsify the proposition, i.e. to fill a knowledge gap about the validity of the proposition. The hypotheses are therefore defeasible in the sense that the falsification of their assertions leads to a retraction and adoption of an initially less likely alternative.

By explicitly reasoning about these knowledge gaps, the system is able to go beyond the current context and knowledge base, addressing the need for reasoning under the open-world assumption. The responsibility for filling those knowledge gaps then falls to external decision-making processes. These processes can then use probabilities to express their confidence in the solutions they provide, thereby addressing the need for capturing the ubiquitous uncertainty stemming from unreliable sensory perception and partial observability of the world.

Acknowledgments. The work reported in this paper has been supported by the EU ICT FP7 projects CogX, "Cognitive Systems that Self-Understand and Self-Extend" (project no. 215181) and NIFTi, "Natural Human-Robot Collaboration in Dynamic Environments" (project no. 247870).

The author would like to thank Geert-Jan Kruijff for insightful and inspiring discussions, and the anonymous reviewers for their detailed and very helpful comments.

References

1. Areces, C.: Logic Engineering: The Case of Description and Hybrid Logics. Ph.D. thesis, University of Amsterdam, the Netherlands (2000)
2. Areces, C., Blackburn, P.: Bringing them all together. Journal of Logic and Computation 11(5), 657–669 (2001)
3. Baldoni, M., Giordano, L., Martelli, A.: A modal extension of logic programming: Modularity, beliefs and hypothetical reasoning. J. Log. Comp. 8(5), 597–635 (1998)
4. Baldoni, M., Giordano, L., Martelli, A., Patti, V.: A modal programming language for representing complex actions. In: Proc. of DYNAMICS'98. pp. 1–15 (1998)
5. Baldridge, J., Kruijff, G.J.M.: Coupling CCG and Hybrid Logic Dependency Semantics. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) (2002)
6. Blackburn, P.: Representation, reasoning, and relational structures: a hybrid logic manifesto. Logic Journal of the IGPL 8(3), 339–625 (2000)
7. Bratman, M.: Intentions, Plans, and Practical Reason. Harvard University Press, Cambridge, MA, USA (1987)
8. Brenner, M., Nebel, B.: Continual planning and acting in dynamic multiagent environments. Journal of Autonomous Agents and Multiagent Systems (2008)
9. Charniak, E., Shimony, S.E.: Probabilistic semantics for cost based abduction. In: AAAI-90 proceedings (1990)
10. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial Intelligence 42, 213–261 (March 1990)
11. Fann, K.T.: Peirce's Theory of Abduction. Mouton, The Hague, The Netherlands (1970)
12. Grice, H.P.: Meaning. The Philosophical Review 66(3), 377–388 (1957)
13. Hobbs, J.R., Stickel, M.E., Appelt, D.E., Martin, P.A.: Interpretation as abduction. Artificial Intelligence 63(1–2), 69–142 (1993)
14. Kamp, H., Reyle, U.: From Discourse to Logic. Kluwer, Dordrecht, The Netherlands (1993)
15. Kate, R.J., Mooney, R.J.: Probabilistic abduction using Markov logic networks. In: Proceedings of the IJCAI-09 Workshop on Plan, Activity, and Intent Recognition (PAIR-09). Pasadena, CA, USA (2009)
16. Kruijff, G.J.M.: A Categorial-Modal Logical Architecture of Informativity: Dependency Grammar Logic & Information Structure. Ph.D. thesis, Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic (2001)

17. Moens, M., Steedman, M.: Temporal ontology and temporal reference. Computational Linguistics 14(2), 15–28 (1988)
18. Poole, D.: Probabilistic Horn abduction and Bayesian networks. Artificial Intelligence 64(1), 81–129 (1993)
19. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall, Upper Saddle River, NJ, USA, second edn. (2003)
20. Schiffer, S.R.: Meaning. Oxford, Clarendon Press (1972)
21. Sgall, P., Hajičová, E., Panevová, J.: The Meaning of the Sentence in Its Semantic and Pragmatic Aspects. Reidel Publishing Company, Dordrecht, The Netherlands and Academia, Prague, Czechoslovakia (1986)
22. Stickel, M.E.: A Prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. Annals of Mathematics and Artificial Intelligence 4, 89–105 (1991)
23. Stone, M., Thomason, R.H.: Context in abductive interpretation. In: Proceedings of EDILOG 2002: 6th workshop on the semantics and pragmatics of dialogue (2002)
24. Stone, M., Thomason, R.H.: 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)