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

DR 6.4: Situated Dialogue with Adapting Levels of Vagueness and Abstraction

Hendrik Zender, Geert-Jan M. Kruijff

DFKI GmbH, Saarbrücken
⟨zender@dfki.de⟩

Due date of deliverable:	July 31 2011
Actual submission date:	July 29 2011
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. For these tasks, the robot was always provided with sufficient a priori dialogue competence to carry them out. This deliverable describes how in Year 3 WP6 explored how such competences could be acquired, as a form of self-extension (Task 6.5), as well as how a robot can verbalize its conceptual knowledge in an adaptive and variable way (Task 6.6). More specifically, we present a revised and extended manuscript on attentional anchor progression in spatially situated dialogues as well as a report on linking commonsense and linguistic knowledge for situated action and interaction. This work is closely related to DR.6.3, where we describe computational approaches to language learning, acquiring communicative competences from the ground up.

Executive Summary

One of the objectives of CogX is self-extension. This requires the robot to be able to actively gather information it can use to learn about the world. One of the sources of such information is dialogue. For this to work, the robot needs to be able to establish with a human some form of mutually agreed-upon understanding, a *common ground*. 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.

The dynamic, interactive setting of CogX in which a robot actively learns requires more than following a fixed, "universal" policy. Learning more, dynamic situations, and the changes in common ground this implies, all require the robot to *adapt* how it acts and interacts, if it is to successfully communicate with a human over time. 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, we let go of this assumption, taking the CogX objective of self-extension to the realm of situated dialogue processing as well. WP6 addressed self-extension for situated dialogue in two aspects. The first aspect concerns the verbalization of large ontologies for modeling common sense indoor knowledge. The categories in these ontologies cover both logical and probabilistic information. Self-extension here focuses on how these categories can be associated with another ontology, namely an ontology of words: WordNet. This provides the means to acquire mappings for verbalizing acquired concepts at varying levels of abstraction and specificity – and in general, it provides a powerful means for large-coverage resources for communicating about indoor environments (Task 6.6). This deliverable, DR.6.4, describes this work in more detail. Furthermore, we continued our work on a second issue in adaptivity, namely varying how much a robot needs to describe to be optimally transparent. We present a revised version of a manuscript that was originally included in DR.6.2. It describes different models for tracking the attentional anchor throughout spatially situated

discourses. The models are evaluated against an empirical production experiment that investigates how humans vary granularity in interaction with a robot, when describing objects in the kinds of small- and large-scale spatial contexts we typically encounter in CogX. The manuscript has been submitted to *Computational Linguistics*, and is currently under review. It makes direct contributions to Task 3.4 "Establishing reference to spatial entities for human-robot interaction" in DR3.2 under WP3.

In contrast, deliverable DR.6.3 takes a further step back, broadening the scope of self-extension in situated dialogue to the aspect of language acquisition per se.

Role of Situated Dialogue in CogX

CogX investigates cognitive systems that self-understand and self-extend. In some 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. Dialogue comprehension and production is ultimately based in a situated, multi-agent model the robot builds up. This model captures epistemic objects like beliefs, intentions and events, in a multi-agent fashion. Such epistemic objects cover both situated and cognitive aspects, and already at this level we see forms of self-understanding and self-extension. In Year 3 we take the notions of self-extension and self-understanding into the domain of situated dialogue itself. We show how the generally prevalent view in CogX on cognition, namely its cross-modal character in understanding the world and deciding how to act in there, naturally extends to the way we can perceive of acquiring competence in situated dialogue processing.

Contribution to the CogX Scenarios and Prototypes

The work of WP6 presented in this deliverable, DR.6.4, contributes directly to the George and Dora scenarios, in relation to work performed in WP3 (Qualitative spatial cognition), WP4 (Planning of action, sensing and learning), WP5 (Interactive continuous learning of cross-modal concepts), and WP7 (Scenario-based integration). The conceptual structures learned by bootstrapping a common-sense ontology of indoor spaces, and their associations with the WordNet ontology are used in the Dora scenario. Further work on quantifying and exploiting such common-sense associations for the tasks of establishing reference to spatial entities for human-robot interaction and planning of efficient robot behavior is reported on in DR3.2 and DR7.2.

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 more information to help it learn more. These efforts always started from the assumption that linguistic resources necessary to talk about the world were available to the robot. The issues WP6 addressed in Years 1 and 2 concerned not so much where these resources were coming from, but rather how they could be used – and how that use could be adapted to optimally fit the context and intentions of the agents involved. The planned work for WP6 in Year 3 is to take the issue of self-extension further, into the domain of situated dialogue processing. In the present deliverable, DR.6.4, we address Task 6.5 and Task 6.6.

Task 6.5 *The goal is to investigate how existing lexico-grammatical knowledge can be extended to cover novel categorical knowledge. We will develop methods for learning two types of mappings: a mapping relating a word's lexical meaning to a predicate-argument structure based on the associations of the category this meaning reflects, and a mapping relating a word's predicate-argument structure to a syntactic family that can express the structure.*

The intention behind Task 6.5 was to see how an existing grammar could be extended with new words, and mappings between those words and meaning representations such that they could be grounded in the knowledge representations of the robot. We have lifted the task from individual words and categories to ontologies of commonsense knowledge, and of words and their categorical organization. As another contribution to Task 6.5, DR6.3 looks at the issue of self-extension from a further step back – by starting to address the problem of how a robot could acquire *linguistic* as well as *commonsense knowledge.*

Task 6.6 *The goal is to provide mechanisms for verbalizing categorical and associative structures of combined concepts. We extend verbalization to cover conceptual structures, focusing on a category and its immediate associations with properties, and other categories. This extends the approach developed earlier in Task 6.1.*

By learning an ontology of indoor concepts and its associations to the WordNet ontology, we have the lexical and taxonomical resources available for verbalizing the robot's spatial knowledge. In order to enable the robot to efficiently, intuitively, and felicitously refer to entities in a spatial domain, we present our work on attentional anchor progression in discourses about large-scale and small-scale spaces. This task also constitutes an intermediate step between Task 4.2 and Task 6.7 in that planning problems can be viewed as conceptual structures that a robot should be able to verbalize.

1.2 Actual work performed

1.2.1 Adaptive abstraction of spatial contexts through anchor-progression in situated discourse

Spatially situated dialogue presents two particular challenges to the field of situated natural language processing. For one, generation of referring expressions (GRE) needs to take into account which information is relevant to successfully identify the referent, while resolving referring expressions (RRE) needs to be aware of the appropriate context within which to uniquely single out the referent. Secondly, viewing the successful establishment of reference from a discourse-oriented perspective involves keeping track of the focus of attention and the resulting context shifts over consecutive utterances. Which entity qualifies as referent and which information is relevant to describe the intended referent is ultimately determined by the previous discourse.

Attentional Anchor Progression in Spatially Situated Discourse. *We present an approach to the task of establishing reference in spatially situated discourses – discourses about domains that need not be confined to the interlocutors' immediate surroundings. We address the particular challenges involved in spatially situated discourse. For one, we present a method for context determination through topological abstraction that can be used for both generating and resolving referring expressions. Secondly, we describe models that predict the progression of the attentional anchor that constitutes the spatial nucleus for context formation. We present an empirical production experiment that evaluates the utility of the proposed methods with respect to situated instruction-giving in tabletop scenes on the one hand, and in indoor environments on the other.*

As an illustration, imagine a service robot that is supposed to clean up an apartment consisting of several rooms. The apartment contains several balls, boxes, and tables (see Figure 1). Rather than expecting an overly verbose instruction like "take the ball in the kitchen and put it into the box on the table in the kitchen", the robot should be able to understand the more natural utterance "take the ball in the kitchen and put it into the box on the table" in the same situation. There might be different boxes that are on tables – rendering the expression "the box on the table" ambiguous with

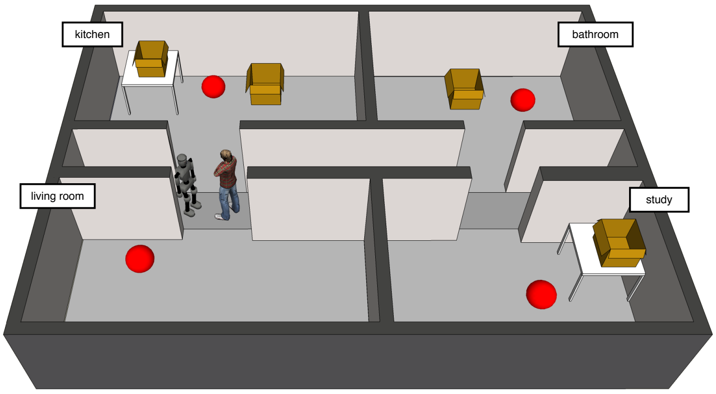

Figure 1: Four room apartment used to illustrate situationally appropriate referring expressions.

respect to the whole apartment. However, the preceding reference to "the ball in the kitchen" has shifted the focus of attention to the kitchen – which in turn allows to felicitously refer to the box on the table in the kitchen as "the box on the table.".

1.2.2 Acquiring a quantitative commonsense ontology for situated action and interaction

In order to plan and act efficiently in partially unknown, large-scale spatial environments, an autonomous robot needs to be equipped with *background knowledge* that provides expectations for where an action might most likely have its intended effect. In particular we are interested in a robot gopher scenario in which the robot has to find household or office objects whose current whereabouts are not known for sure. Moreover, having in mind an intelligent and interactive robot gopher, the robot must also be endowed with the linguistic capabilities to talk about its environment.

Linking Commonsense and Linguistic Knowledge for Situated Action and Interaction. *We present an approach to endowing a mobile indoor robot with a quantitative ontology of indoor commonsense knowledge that can be used for automated reasoning and planning, as well as for processing of situated human-robot dialogue. We leverage existing resources of commonsense and lexical-semantic knowledge: the Open Mind Indoor Common Sense project and the WordNet database, respectively. We describe how frequency estimates obtained from queries to an online image search engine help quantifying commonsense knowledge. This quantified commonsense knowledge is used as informative priors in automated reasoning and planning for efficient*

robot behavior. We furthermore discuss how taxonomical structuring over word senses can be applied to a flat, word-based large-scale commonsense knowledge repository, yielding a rich, quantified ontology of objects and their typical indoor locations. Finally, we describe how the combined linguistic-conceptual commonsense ontology can be used to automatically extend the lexicon of the robot's dialogue system.

The Open Mind Indoor Common Sense (OMICS) project by Honda Research Institute USA Inc. is an effort towards collecting commonsense knowledge from many people with the intention to use it for intelligent robots. One part of it is a collection of user contributed facts about which objects can be found where in typical indoor environments, but there is no explicit quantification of the individual, different likelihoods.

For instance, that collection lists the following locations as typical places for finding a sink: kitchen, laundry, washroom, bathroom, restroom, garage, bar, and laundry room. This would prevent the robot from starting to look for a sink in the living room, but it does not allow the robot to make judgments about whether it should start searching for a sink in the bathroom or in the garage. We therefore developed a method for obtaining quantitative priors for the likelihood of an observation of the respective object in a given set of rooms. These priors allow a decision theoretic planner to decide about the prioritization of searching different places. Table 1 shows some examples of the obtained cooccurrence priors.

Table 1: Cooccurrence matrix for some select objects and locations.

	kitchen	bathroom	garage	office
sink	0.394958	0.24747899	0.053361345	0.05630252
faucet	0.45874125	0.40419582	0.018181818	0.043776225
computer	0.048387095	0.028830646	0.019112904	0.111693546

In order to overcome problems with ambiguities in natural language, as well as to ensure that the lexicon of our dialogue system covers the kinds of things the robot knows about, we combine the OMICS-derived cooccurrence matrix with the WordNet lexical database. For instance, OMICS lists, among others, office, computer lab, and music room as likely places for keyboards. To disambiguate between the musical instrument and the computer input device, we manually annotated the OMICS assertions with WordNet word senses. On top of that we can then construct an ontology that is based on the taxonomical structure represented by the WordNet hypernym/hyponym relations. Finally, for all distinct words we create lexical entries in our CCG-based grammar to allow the robot to communicate about its environment in natural language.

1.3 Relation to self-extension

The presented work fits well with the goal of devising mechanisms for self-extension. Notions that we consider central for self-understanding and self-extension are *relevance* and *context*. In order to be able to take action towards knowledge extension, the robot must know which part of its knowledge and which part of the situation of its environment – *context* – give rise to the current need to self-extend. At the same time, having diverse knowledge sources at its disposal (e.g., gathering new evidence through observing or interacting with the environment, or asking a human, or consulting large-scale knowledge repositories), the robot must be able to make an informed decision about which knowledge to incorporate (and also which action to take in order to be able to obtain such knowledge) – *relevance*.

The models for extending linguistic knowledge and for spatially situated dialogue presented in this deliverable allow the robot to successfully refer to things in its immediate and distant environment alike. This is a crucial prerequisite for a robot that is supposed to extend and validate its knowledge through situated dialogue with a human. They provide the robot with a notion of context.

The ontological, quantitative models of commonsense knowledge, on the other hand, support self-extension with a notion of relevance. They present informative priors for efficient planning under uncertainty, allowing the robot to focus on prioritizing actions that are more likely to yield the desired (epistemic) effects. Secondly, the models provide more structure than flat, unstructured collections of isolated facts. Assessing the relational structure of the vast collection of knowledge obtained by combining large-scale online repositories of commonsense with lexical-semantic knowledge it becomes possible to decide on a relevant subset of new knowledge to be internalized – precluding an explosion of search space through largely irrelevant knowledge.

Together, this endows the robot with *focus* for making economic use of its sparse computational resources as well as for minimizing the time it takes the robot to discover necessary knowledge by exploring its environment.

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.

Whereas most approaches to the task of generating referring expressions (REs) consider small visual scenes (such as tabletop or single-room domains), Paraboni *et al.* [3] are among the few to address the issue of generating referring expressions to entities outside the immediate environment. They present an algorithm for context determination in hierarchically ordered domains. We present an extension of our previous work on determining appropriate contexts in spatial domains [5, 6]. We advance the state-of-the-art by not only looking at single REs, but rather taking into account how the focus of attention shifts through the spatial domain during a discourse. The approach lends itself to be used with the kinds of spatial knowledge bases that are investigated and developed in WP3 (cf. DR.3.2).

There exist several large-scale collections of commonsense knowledge, e.g., OpenCyc[1], the Open Mind Common Sense[2] project, or WordNet[3]. They provide vast resources of knowledge about the world with the explicit aim of enabling machines to "understand" the world. Their size is also one of the factors that aggravate their use in intelligent robots. A robot does not necessarily have to be equipped with complete knowledge about the world. It arguably only needs to possess the knowledge that enables it to achieve its current tasks (more efficiently). The second issue with these commonsense knowledge databases is that they must be resolvable to the robot's own observations and actions in order to be truly meaningful and useful. We leverage existing state-of-the-art resources of commonsense and linguistic knowledge in order to endow our robot with a more complete representation of the concepts in its intended domain of operation, as well as with the means to talk about its (abstract as well as concrete) knowledge about its environment.

[1] http://www.opencyc.org/
[2] http://openmind.media.mit.edu/
[3] http://wordnet.princeton.edu/

2 Annexes

2.1 Zender and Kruijff, "Attentional Anchor Progression in Spatially Situated Discourse"

Bibliography H. Zender and G.J.M. Kruijff. "Attentional Anchor Progression in Spatially Situated Discourse" Submitted to *Computational Linguistics.*

Abstract We present an approach to the task of generating and resolving referring expressions in spatially situated discourses – discourses about domains that need not be confined to the interlocutors' immediate surroundings. Spatially situated communication requires the interlocutors to draw attention to entities that are not currently observable as well as to understand which entities in the world are being talked about. As such a discourse unfolds, the focus of attention shifts, introducing new entities and making new situations available for reference. Spatially situated discourse thus presents two particular challenges to the field of situated natural language processing. For one, generation of referring expressions (GRE) needs to take into account which information is relevant to successfully identify the referent, while resolving referring expressions (RRE) needs to be aware of the appropriate context within which to uniquely single out the referent. Secondly, viewing the successful establishment of reference from a discourse-oriented perspective involves keeping track of the focus of attention and the resulting context shifts over consecutive utterances.

We address these challenges with a method for context determination through topological abstraction that can be used for both generating and resolving referring expressions, and with models that predict the progression of the attentional anchor that constitutes the spatial nucleus for context formation. We present an empirical production experiment that evaluates the utility of the proposed methods with respect to situated instruction-giving in tabletop scenes on the one hand, and in indoor environments on the other.

Relation to WP This revised manuscript presents a more detailed discussion of the models for tracking the attentional anchor presented earlier [4]. The models are suitable for the tasks of generating and resolving referring expressions alike. The report describes how the models can be directly and straightforwardly used with the dialogue framework developed in the context of WP6 and the models for representing large-scale space from WP3 (see also DR6.2). The results of this work hence contribute immediately to the Dora demonstrator from WP7 because they allow a robot to produce and understand consecutive references to entities that are located elsewhere in its operating environment (such as an office floor or an apartment).

2.2 Zender and Kruijff, "Linking Commonsense and Linguistic Knowledge for Situated Action and Interaction" (Report)

Bibliography H. Zender and G.J.M. Kruijff. "Linking Commonsense and Linguistic Knowledge for Situated Action and Interaction" Technical report.

Abstract In this paper, we present an approach to endowing a mobile indoor robot with a quantitative ontology of indoor commonsense knowledge that can be used for automated reasoning and planning, as well as for processing of situated human-robot dialogue.

We leverage existing resources of commonsense and lexical-semantic knowledge: the Open Mind Indoor Common Sense project and the WordNet database, respectively. We describe how frequency estimates obtained from queries to an online image search engine help quantifying commonsense knowledge. This quantified commonsense knowledge is used as informative priors in automated reasoning and planning for efficient robot behavior. We furthermore discuss how taxonomical structuring over word senses can be applied to a flat, word-based large-scale commonsense knowledge repository, yielding a rich, quantified ontology of objects and their typical indoor locations. Finally, we describe how the combined linguistic-conceptual commonsense ontology can be used to automatically extend the lexicon of the robot's dialogue system.

Relation to WP This report presents an approach for creating an ontology that combines quantitative commonsense knowledge with lexical-taxonomical knowledge. It can readily be used to extend the lexicon of the CCG-based natural-language dialogue framework under development in WP6. It also presents an ontological knowledge base for the conceptual spatial representation in WP3. The constructed cooccurrence priors additionally serve as a knowledge base for the switching planner in WP4. The reported work also has an immediate impact on the Dora demonstrator from WP7.

References

[1] Michael 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.

[2] Moritz Göbelbecker, Thomas Keller, Patrick Eyerich, Michael Brenner, and Bernhard 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.

[3] Ivandré Paraboni, Kees van Deemter, and Judith Masthoff. Generating referring expressions: Making referents easy to identify. *Computational Linguistics*, 33(2):229–254, June 2007.

[4] Hendrik Zender, Christopher Koppermann, Fai Greeve, and Geert-Jan M. Kruijff. Anchor-progression in spatially situated discourse: a production experiment. In *Proceedings of the Sixth International Natural Language Generation Conference (INLG 2010)*, pages 209–213, Trim, Co. Meath, Ireland, July 2010. Association for Computational Linguistics.

[5] Hendrik Zender, Geert-Jan M. Kruijff, and Ivana Kruijff-Korbayová. A situated context model for resolution and generation of referring expressions. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009)*, pages 126–129, Athens, Greece, March 2009. Association for Computational Linguistics.

[6] Hendrik Zender, Geert-Jan M. Kruijff, and Ivana 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.