

From Sensors to Sense: Integrated Heterogeneous Ontologies for Natural Language Generation

Mihai Pomarlan and Robert Porzel and John Bateman and Rainer Malaka*

University of Bremen

{pomarlan, bateman}@uni-bremen.de {porzel, malaka}@tzi.de

Abstract

We propose the combination of a robotics ontology (KnowRob) with a linguistically motivated one (GUM) under the upper ontology DUL. We use the DUL Event, Situation, Description pattern to formalize reasoning techniques to convert between a robot’s beliefstate and its linguistic utterances. We plan to employ these techniques to equip robots with a *reason-aloud* ability, through which they can explain their actions as they perform them, in natural language, at a level of granularity appropriate to the user, their query and the context at hand.

1 Introduction

It is a sunny afternoon in the not too distant future, and Elroy wants to play ball in the garden with Rosie the robot. He finds her moving about in the dining room and asks “What are you doing?”. “I am busy”, Rosie answers, politely but suggesting she doesn’t want to be interrupted right now. Disappointed, but not wanting to let go just yet, Elroy presses on. “What are you doing?” he asks again. “I am setting the table,” Rosie answers. Still not satisfied he repeats his question again and Rosie explains “I am bringing cutlery and plates to the table and currently looking in this cupboard for a spoon and fork for Judy. They must not be plastic, for she is allergic to it.”

The little scene above shows an interaction between a human and a household robot where the appropriate level of granularity with which the

robot should describe its task varies greatly as the dialog situation evolves. Generally, such interactions cannot be restricted to command-giving (by the human) and command-taking (by the robot). Even a specialized device, e.g. a coffee machine, offers some feedback about its state. Indeed, the spectrum of possible interactions can be quite complex: the robot might ask for a way around an obstacle it encountered in a task, discuss user preferences and task schedules, take initiative in asking for parameters of upcoming tasks, or ask the users about their activities, as these will affect the robot’s task planning and execution.

Compared to more complex situations, the one in our example scene seems simple, but it nevertheless captures an aspect that will be important for the interlocutory capabilities of robots: the ability to interpret events and to describe them understandably, at a level of granularity appropriate for the user and their query. This requires integrating heterogeneous forms of knowledge, such as records of sensor data, representations of activities at different abstraction levels, and theories about the environment and the interlocutory partners.

For this undertaking, we envision a *reason-aloud* capability for robotic agents, analogous to human *think-aloud*. Humans are quite capable of reflecting overtly on their actions and describing them in parallel to their execution, which is why the *think aloud protocol* has become widely used in numerous studies in cognitive science, psychology and human-computer interaction (van Someren and Barnard, 1994). For this a situated artificial agent must combine knowledge of the activities at hand with the knowledge required to express them declaratively.

This work was partially funded by Deutsche Forschungsgemeinschaft (DFG) through the Collaborative Research Center 1320, EASE.

2 Approach

Our approach is to extend an Ontology for Everyday Activities, originally developed as part of the EASE project in robotics (Beetz et al., 2018). We base this extended ontology on the principles proposed by Masolo *et al.* using the DOLCE+DnS Ultralite ontology (DUL) as an overarching foundational framework (Masolo et al., 2003; Mascardi et al., 2010). The purpose of the ontology is to extend the KnowRob ontology to support more natural, commonsense interactions concerning everyday activities in robotics. Specific branches of the KnowRob knowledge model pertaining to everyday activities (Beetz et al., 2018), such as those involved in table setting, have already been aligned to the DUL framework. Additional axiomatization that is beyond the scope of description logics is integrated by means of the Distributed Ontology Language (Mossakowski, 2016). The extension we consider in this paper is for adding language generation capabilities, to which end we align the linguistically motivated ontology GUM (Bateman et al., 2010), and its extension to spatial concepts, to DUL and the EASE ontology.

The key advantage of this ontological alignment via DUL is first and foremost a bridge between the KnowRob system, a mature knowledge processing system for robotics (see section 3) and language generation software that uses GUM representations, such as KPML (Bateman, 1997). Using the DUL-specific *Descriptions and Situations* pattern, we can employ these to supply concepts and reasoning methods for the problem of interpreting `Events` into `Situations` and constructing `Descriptions` for them (see section 4.1).

We will only look at command-taking and the robot performing a “reasoning aloud” (analogous to human “think aloud”) in this paper. We hope the reasoning techniques enabled by our approach will lay a scalable base for future work on more complex interactions, e.g. dialogical negotiating when activities conflict, but we stress that a “reasoning aloud” capability can be useful on its own. It shows understanding on the robot’s part of the task it performs, and makes the robot itself more understandable to the user.

3 KnowRob and KPML

KnowRob (Beetz et al., 2018) is a software system to integrate and reason with a variety of robotics knowledge sources. Its interface is a database

query system via Prolog predicates, providing a uniform way to access the reasoning mechanisms underneath. These mechanisms can, however, be varied by employing an approach called *computables* which allows for predicates to map to and take results from functions appropriate for a task.

In this way, KnowRob can do hybrid reasoning on symbolic data - which it queries or infers from a logical database - as well as raw data - such as sensor readings and log files. Reasoning mechanisms can make use of logical axioms, but also perform collision or visibility testing in an environment and draw on inverse kinematics, physical simulation, etc. To handle uncertainty, KnowRob uses probabilistic, first-order relational models. These models are intended to capture general principles about similar objects. For example, they may represent a probability distribution on where to look for an item, or where to store it in a kitchen, given its type.

To handle environment dynamics, the KnowRob ontology includes some concepts for Actions and their Effects. We have extended the ontology’s coverage in this respect and brought it into alignment with DUL. Also, the KnowRob ontology defines concepts that have been used to construct what are termed within the EASE project as NEEMs (Narratively Enabled Episodic Memories), which are comprehensive records of a robot’s activity: this includes what the robot has observed through its sensors, how it acted in the world, its task tree (from which a hierarchy of *intentions* is discernible) and the execution status of tasks. KnowRob contains predicates to select and reason with `Events` recorded in the NEEMs, including temporal calculi. NEEMs were intended as data collection for learning, to improve robot performance. Expert users can employ them to debug the robot. On their own however, they are too large and incomprehensible for the average user to handle, making natural language techniques highly relevant.

For generating comprehensible and appropriate language we propose to employ KPML. This system offers a well-tested platform for grammar engineering that is specifically designed for natural language generation (Reiter and Dale, 2000). KPML employs the use of large-scale grammars written with the framework of Systemic-Functional Linguistics (SFL). The employment of SFL enables us to include linguistic phenomena

which are important for the generation of natural texts alongside the propositional content that is to be expressed (Bateman, 1997).

In the following, we will outline how the respective interleaving of the symbolic layers of KnowRob and the ontological model of GUM via DUL facilitates crossing the bridge from a robot executing particular actions to talking about them in real time. As stated before, we also are working on using the same bridge to enable the robot to understand linguistic input, i.e. instructions.

4 From Language to Beliefstate– and back again

4.1 Event, Situation, Description

We will first summarize a few DUL concepts that are central to our approach. *Events* are either *Processes* or *States*, in which several objects may participate. An *Event* is related to one or more *Situations*, which are *views* on (or interpretations of) an *Event*. A *Situation* satisfies, or is consistent with, a *Description*. As an example, a robot’s movements and the contacts between objects that they cause would be events. A situation would be the robot executing a plan for table setting. The table setting plan itself would be the description consistent with the situation.

A robot’s knowledge cuts across all these distinctions. The robot causes, observes, and records events as they happen. It may be situated as executing a task, or interacting with a user towards some purpose. And it has theories of the environment around itself, e.g. action, environment, and user models, as well as higher-level plans.

Most generally, communication between user and robot involves the two exchanging descriptions, for which we identify two problems:

- command/inform: the robot receives a linguistic description. It creates new descriptions and situations as appropriate so as to update its belief state about the world or begin executing a requested task.
- reason aloud: the robot has a record of events, a representation of the situations it is in, and various descriptions. It summarizes this knowledge into a description, to answer a query at an appropriate level of granularity, without overwhelming the user.

The purpose of our combined ontology is to enable reasoning techniques to bridge these conver-

sions: events to situations, and situations to descriptions. All the more specific components are consequently related to the DUL backbone.

4.2 Events ↔ Situations

The direction especially relevant for us here is going from events to situations that interpret them. The opposite, from situations to events, means simply that the robot causes events in the world according to some chosen plan. For this purpose, we define several classes of situations in our ontology, with restrictions to specify when it is appropriate to use the situation as an interpretation for the set of events. Several situations may be appropriate to interpret a set of events. Situations include:

- an agent (human/robot) acting on inanimate objects, e.g. ‘Actor Creates Something’, ‘Actor Affects Something’, ‘Resource Absent’.
- human-robot interaction, e.g. ‘Command Issued’, ‘Availability Query’.
- inanimate objects acting on each other, e.g. ‘Stable Placement’, ‘Physical Interaction’.

Usually, choosing an interpretation when the robot is the only active agent in the events is straightforward; the robot “knows” what its task tree is, i.e., what it wants to do, because for the robotic system we use the programs it runs are semantically annotated with goals.

Finding an appropriate situation in other cases either implies guessing the other agent’s intentions, for which probabilistic reasoning or simulation can be used to find the most likely intentions given the observed evidence, or, if there is no active agent in the event, parsing an event timeline according to a grammar of situations (cf. (Beßler et al., 2018b) for an action parser using the DUL and KnowRob ontologies).

4.3 Situations ↔ Descriptions

We will first look at describing a situation to the user. Some situation classes in our ontology have unique description correspondents, e.g., “Actor Creates Something” has GUM’s “CreativeMaterialAction”, while others may define, via restrictions, subsets of descriptions applicable to them.

To construct description individuals – filling in semantic roles – we use a method employed in KnowRob for assembly planning (Beßler et al., 2018a) which checks that an individual asserted

to belong to a class actually respects restrictions placed on that class, in particular whether it is linked to other individuals by appropriate object properties. If this is not the case, the method creates new individuals and relations as needed. Restrictions on fillers for a description's semantic frame roles can be written in SWRL.

We will also investigate reasoning methods to update the interaction situation in the robot's beliefstate based on user utterances. These will be semantically analyzed and interpreted as commands or queries. For commands, robot programs will be constructed using blocks from a library of basic actions. Query answering involves the event-situation-description bridges described previously.

As an example of how our approach is intended to work, consider the following scenario: the robot has "setting the table" as its top-level task, and it knows this task is intended to prepare another task ("eating") to be done by other agents. The current subtask the robot is performing is "picking" a spoon. Note, mechanisms to represent and reason about task trees are already in place in our knowledge processing system.

Suppose the robot decides to report that it is "setting the table", which is a particular type of situation captured by a broad situation concept AgentAffectsSomething. Our ontological characterization is that a AgentAffectsSomething individual satisfies some gum-DispositiveMaterialAction, so we create an individual of this latter type to describe what the robot is doing.

Individuals of type gum-DispositiveMaterialAction should obey certain restrictions however. One such restriction is such an individual should have an actor that is some GUMThing, and our newly created individual has no such information attached yet. To enforce this restriction, an agenda item is generated to create and look for a suitable actor, which in this case will be a description of the agent of the "setting the table" situation.

Where needed one can go beyond restrictions placed on descriptions in the GUM. For example, suppose we want the robot to say why it is "setting the table". In this case, we add a new restriction on the newly created gum-DispositiveMaterialAction individual, that it should have as reason some GUMThing, and this will result in an agenda item to look for a filler for this role, which will be a description of the task that "setting the table" prepares.

What the user should be told as part of a "think-aloud" protocol depends on what the robot thinks the user might know about the robot's task, so let's suppose as an example the user knows nothing. The question then is what to report from the task tree, which will probably have very many nodes? Several heuristics may be tried here, but they can be formulated in terms of the task tree structure. One such heuristic is to report the current subtask, "picking", the robot's top-level task, "setting the table", and the task being prepared by the robot's top-level task, "eating".

Each of these situations gets a Description individual of appropriate GUM type. There is flexibility in choosing which of the three gets to be the main clause of the resulting utterance and which get to be dependents, which offers us flexibility in generating a report:

I'm picking up the spoon because I'm setting the table so people can eat.
I'm setting the table because people will eat, therefore I'm picking up the spoon.
People will eat soon therefore I'm setting the table so I'm picking up the spoon.

4.4 Matching the Description Granularity

There may be several parses of a set of events, several situations that are possible views on them, and several descriptions for each situation; e.g., levels of abstraction at which to report in the reasoning aloud. Fortunately, the graphs representing the situations already feature different levels of generality. For example, a situation where we encounter a "grasp - lift - place - release" pattern will be categorized as a "pick and place" action, which, in turn, can be part of a more general activity such as "table setting". The hierarchy and the respective distances in the graph has to be aligned to the information stemming from the interaction situation to pick out which level of abstraction to report.

Numerous approaches have been proposed to control such alignments. Very prominent in natural language generation are approaches based on user modeling, e.g. the TAILOR system (Paris, 1988). However, also discourse modeling (Pfleger et al., 2003) or the situational context (Porzel, 2009) come into play when selecting the propositional level of granularity. Formally, levels of granularity can also be expressed as a set of theories forming a hierarchical structure (Hobbs, 1985). Nevertheless, a concrete method for matching these structures has to be found and tested.

References

- John Bateman. 1997. *Enabling technology for multilingual natural language generation: the KPML development environment*, volume 3(1):15–55. *Journal of Natural Language Engineering*.
- John A. Bateman, Joana Hois, Robert J. Ross, and Thora Tenbrink. 2010. [A linguistic ontology of space for natural language processing](#). *Artificial Intelligence*, 174(14):1027–1071.
- Michael Beetz, Daniel Beßler, Andrei Haidu, Mihai Pomarlan, Asil Kaan Bozcuoglu, and Georg Bartels. 2018. Knowrob 2.0 – a 2nd generation knowledge processing framework for cognition-enabled robotic agents. In *International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia.
- Daniel Beßler, Mihai Pomarlan, and Michael Beetz. 2018a. Owl-enabled assembly planning for robotic agents. In *Proceedings of the 2018 International Conference on Autonomous Agents, AAMAS '18*, Stockholm, Sweden. Finalist for the Best Robotics Paper Award.
- Daniel Beßler, Robert Porzel, Mihai Pomarlan, Hagen Langer, John Bateman, Rainer Malaka, and Michael Beetz. 2018b. Foundational models for manipulation activity parsing. In *Proceedings of the 2018 International Conference on Robotics and Automation (ICRA) (submitted for review)*.
- Jerry R. Hobbs. 1985. [Granularity](#). In *Proceedings of the 9th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'85*, pages 432–435, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Viviana Mascardi, Valentina Cord, and Paolo Rosso. 2010. Technical report disi-tr-06-21, university of genua.
- C Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, and A Oltramari. 2003. Wonderweb deliverable d18 ontology library.
- Till Mossakowski. 2016. [The distributed ontology, model and specification language - DOL](#). In *Recent Trends in Algebraic Development Techniques - 23rd IFIP WG 1.3 International Workshop, WADT 2016, Gregynog, UK, September 21-24, 2016, Revised Selected Papers*, pages 5–10.
- Cécile L. Paris. 1988. [Tailoring object descriptions to a user's level of expertise](#). *Comput. Linguist.*, 14(3):64–78.
- Norbert Pflieger, Jan Alexandersson, and Tilman Becker. 2003. A robust and generic discourse model for multimodal dialogue. In *Proceedings of the 3rd Workshop on Knowledge and Reasoning In Practical Dialogue Systems*.
- Robert Porzel. 2009. *Contextual computing for natural language processing*. Ph.D. thesis, University of Bremen.
- Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press, Cambridge, U.K.
- Maarten van Someren and Yvonne Barnard. 1994. *THE THINK ALOUD METHOD: A practical guide to modelling cognitive processes*. Academic Press, London, U.K.