Task-based Evaluation for Improving Natural Language Understanding in Human-Robot Rescue Teams

Fereshta Yazdani
Institute for Artificial Intelligence
University of Bremen
Bremen, Germany 28359
Email: yazdani@cs.uni-bremen.de

Matthias Scheutz
HRI Laboratory
Tufts University
Medford, MA 02155, USA
Email: matthias.scheutz@tufts.edu

Michael Beetz
Institute for Artificial Intelligence
University of Bremen
Bremen, Germany 28359
Email: beetz@cs.uni-bremen.de

***Abstract*—Mixed human-robot teams are increasingly considered for accomplishing complex mission due to their complementary capabilities. A major barrier for deploying such heterogeneous teams in real-world settings, is the current lack of natural skills in robotic team members, such as the understanding and interpretation of natural language instructions that include referential descriptions of entities in the world. In this paper we report the results of an empirical study in which humans tend to use referring expressions. We show how the received results and ideas can be used as guidelines to improve dialogue systems. By integrating and extending our system with these results, we will show how complex natural language instructions can be easily translated by robotic systems.**

I. Introduction

Due to the continuely rising number of applications for human-robot teaming, there is an increasing need for technologies based on natural human-robot interaction (HRI) in jointly accomplishing tasks in such teaming scenarios. An example shows the project Sherpa [9] that investigates how a human team leader commands a team of robots to perform part of a search task in a mountain environment. An exemplified command might be *"Go behind that tree. It is the tallest"*. This command includes linguistic as well as visual descriptions of the terrain that have to be interpreted by the robots in the context of the task. Enabling robots to understand such natural language (NL) instructions would facilitate seamless the coordination in human-robot teams, however interpreting those kind of instructions is still a challenge in particularly if the robot has no prior knowledge of the constructs applied in the language. Misinterpretations by robots can lead to wrong results and disastrous consequences in search and rescue missions, especially, when finding injured persons in dangerous areas is very time limited and has to be done quickly. To accomplish such kind of instructions, the robotic agents must be able to understand verbal references to physical entities perceived in the world.

Our research focuses on making human-robot teams more effective for search and rescue missions by:
(1) ensuring natural taskability of the human operator and his robotic team. When working with robot teammates NL descriptions such as reference resolution, perceptual and spatial descriptions need to be resolved in order to assure correct task interpretations.
(2) understanding intentions and task-based instructions of the human team leader. While working with humans, robots need to display their intentions and interpretations of instructions in such a way that are easily readable by the human team leader to ensure correct task interpretation.
(3) geometric and cognition-enabled reasoning on tasks for generating and refining goals. In order to execute tasks correctly, robots request in form of prolog queries knowledge about context-specific tasks that are replied by a map with the appropriate knowledge.

Our investigation uses interaction scenarios from human-robot visual search and rescue tasks in simulation. The human commands the robot using NL, and the robot, in turn, has to interpret those commands in the context of the task. This includes make inferences about where and how it has to position itself to have unrestricted perceptual access to target locations in the environment. In human-robot team settings robots need to have an understanding of human expectations about communication. The goal is to make those interactions as natural as possible. So, we encourage this work by creating hypotheses on human instructions to improve dialogue systems and NL understanding for robots.
We hypothesize that **(H1)** humans focus on *prominent entities* in the world while giving instructions. While humans guide other humans through the terrain, they use specific descriptions based on the shape or the size of the perceived entity, such as "small", "big", etc. They also use spatial relations to entities how they are related to each other.
In addition, we hypothesize that **(H2)** humans do not explicitly mention obstacles in the environment when routing another human. They expect that their team is able to perceive these changes and obstacles in the terrain and to overcome them on their own.
To be able to develop the necessary NL understanding and reasoning mechanisms, we conducted an empirical study of the visual search and rescue task and checked these hypotheses. We will report the results together with an analysis that can be used to improve current NL understanding and dialogue systems. We have extended our task interpretation system with components of natural language processing and integrated those ideas into that system to ensure the interpretation from

high-level NL commands into low-level action plans that are comprehensible for robots.
The rest of the paper is organized as follows: We start with a review of existing work in HRI and then introduce briefly the experimental setup of the empirical study. Next, we present the results and their significance by integrating them into our task interpretation system that uses the concept of interpreting vague tasks through sampling- and simulation-based reasoning. At the end we discuss our finding results and future directions.

II. Related Work

The HRI field is of emerging and increasing interest. Much of the research effort has focused on developing computational models of social intelligence for robots to allow them to successfully interact with humans as assistants or teammates in a natural and intuitive way [5], [7].
In [8] is discussed that humans are performing task planning, monitoring and supervision while robots are acting as intelligent, autonomous assistants and interacting symbolically and physically. This interaction is done via NL, gestures and touch.
An essential interest in HRI is making robots more like human beings especially many researchers have been developing robots that perform human tasks [4], [11], [12]. A lot of work was done in the area of telepresence where humans remote robots to explore remote environment or to interact to other people between long distance. The idea of employing robots as partners to work with human together or to replace humans as team member to be outside the rubble is of huge interest in many fields starting from virtual games towards space exploration and rescue missions [3], [14], [16].
Another essential field in HRI is the use of NL interactions during task execution between humans and robots. Krause et al. [6] use NL as a way of one-shot learning of visual objects and enables the robot to immediately recognize the described object. Further work based on NL and resolving references to visual objects is done in [15] where an algorithm for reference resolution is presented that identifies the referents of referential expressions and modifies the world model based on such expressions.

It is known that HRI studies are explicitly organized and concise e.g. what the human is saying and the robot is performing. Also tasks are very well and clearly structured, so the spectrum of received data is kept very small. Most of these studies are carried out in indoor environments where the search space is limited and the different target objects are clearly arranged, thus allowing the human to correct the robot if it misinterprets a task description. The situation is, however, different for outdoor terrains where terrain complexity and task conditions may not allow such human intervention. In this paper we focus on how instructions are formulated for outdoor environments, i.e., which types of route and referential descriptions with different target objects humans will use that need to be processed by the robot to successfully interpret the human commands. Our idea is to understand the intention, the language and the behavior of humans when tasking other humans to achieve common goals and apply that on robots. We will present the empirical experiment we conducted to collect data from human instructors about how instructions would be given for outdoor search tasks.

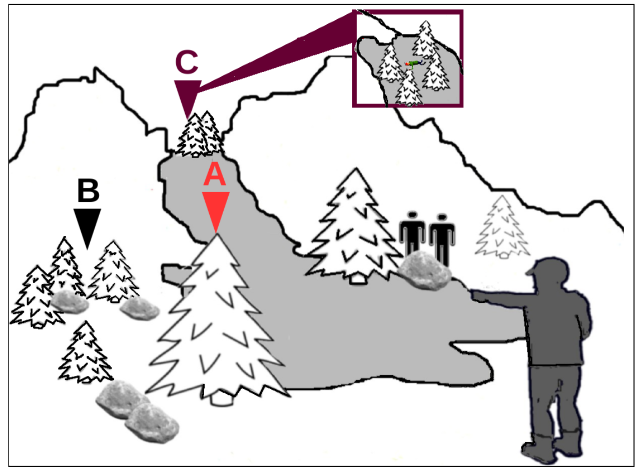

Fig. 1. An illustration of a mountain environment with a human team leader and his team members, three landmarks marked with arrows to point at locations of interest.

III. The Search And Rescue Experiment

The main goal of the online experiment using Amazon Mechanical Turk was to investigate the use of referential expressions used to describe task-relevant real-world entities in outdoor settings. The experiment was done with human teams to investigate how humans command other humans in a natural way. The idea is to convey this kind of communication to robot teams to enable natural tasking.

Methods: We created a virtual mountain environment and took several images showing outdoor scence to specify tasks to be accomplished by the participants. These images include possible landmarks that present locations of interests where injured persons are detected (e.g., see Figure 1 for an example). These indicated landmarks were shown to the participants to mark the positions where their team has to go.

The figure shows a mountain terrain with an avalanche between the mountains, surrounded by many trees and rocks. In the right corner of this figure is a human team leader who has an overview of the terrain and is giving instructions which have to be accomplished by the team. These instructions are based on the given landmarks indicated with arrows. Each landmark represents a task which the participants had to perform by giving instructions to their team starting at their initialpose.
The overaching goal of the mission was to find injured persons trapped or hidden by entities in the terrain. So, the tasks involve sending the team to find injured persons at the locations indicated by the arrows while considering the conditions of the environment such as obstacles. Participants then had to give written instructions using the keyboard to their team as “team leader” that would guide them through the terrain to the specific landmark to find injured persons. Subjects were informed that their teams did not have any knowledge about the terrain and its conditions, so the participants had to consider these while instructing. They had to formulate their instructions clear and comprehensible to the team members to avoid misinterpretations and ambiguities. In order to keep the focus on NL, we avoided to use gestures that could facilitate the interpretation e.g. by pointing into specific directions and placed the given landmarks next to specific entities in the

terrain. The study had no time limitation to allow participants sufficient time to accomplish the tasks and to receive efficient and effective results at the end. We integrated the hypotheses (see Section I) into the study by modifying the terrain explicitly. The *H1* was integrated by changing the shape and size of some entities in the world that were close to a given landmark (see Figure 1, Landmark A). The *H2* was used by placing obstacles in the images e.g. an avalanche along the way. By integrating those hypotheses in the tasks, we wanted to know if the participants consider those as significant information and share them while commanding their team.

Participants: In the study, 69 participants provided effective contributions (44 male, 25 female) with an average age of 30.33 years. All participants were English native speakers. The study was conducted with non-experts to receive a huge spectrum of natural and various sentences.

IV. Data Analysis and Discussion

To help improving the effectiveness of current NL processing algorithms for robots, we next briefly analyze and evaluate the data based on the given hypotheses to be able to provide empirically grounded guidelines. We specifically investigated the ways in which the participants used language to encode relations between entities, places and other aspects of the environment to convey knowledge or to resolve ambiguity. The various examples given in this section are contributions of the participants in the study.

A. Simple Instructions

The results of our study show that the several instructions are often composed of the same basic structure

*<**V**ERB><**O**RDER><**D**ESCRIPTION>*

The *VERB* expresses a physical action in the sentence that describes what the team has physically to do such as *going* and *walking*. In combination with the *ORDER* the team is commanded to do something. An example might be *Go right of tree!* that describes an action with "Go", an order "right of" and a description "tree" in an outdoor mission. Together with the *DESCRIPTION* an object can be denoted and described in sufficient detail that the team can form for instance a mental picture or an understanding of it.

We defined this simple *VOD-structure* to generate rules for syntactically and semantically interpretation of instructions into robotic action plans. For the interpretation into action plans, we have used a system that was previously introduced in [17] and extended it by components for command interpretation to enable human-robot interaction in search and rescue missions. In the next section we will show an experimental application by introducing and evaluating the expanded system with the results of the study.

Depending on the various tasks and the level of detail in the contributions of this study, this *VOD-structure* was extended by various constructs in the results that will be presented in the remainder of this section.

B. Action Descriptions vs. State Descriptions

The instructions generated in the study can be divided into two types: *action descriptions* and *state descriptions*. Action descriptions might be a listing of actions whereas state descriptions might specify conditions of perceivable entities. An example of action descriptions is given with

(b) Move to your left towards me to get free of the trees. Then move left skirting the avalanche until you reach a hill. Move up the hill laterally and forward towards the mountains until you reach the site.

It shows that the actions include pre- and postconditions dependent on each other. They have to be considered accordingly during the task execution as emphasized by the preposition *then* that has been used to indicate what happens next. The conjunction *and* presents a sequential arrangement of the actions that shows the condition of a successive action execution. Other prepositions have been used to emphasize instructions in the manner of time and place.

The state descriptions specify conditions of the environment that also include perceivable entities. The instruction

(c) You are heading towards the tallest tree in the set of trees in front of you. It is the last one in the group and near some large rocks

that was given in the results shows a sequence of visual descriptions of the environment. Entities that are distinctive described using scalar terms such as *tallest tree* or *large rocks*.

C. Referential Descriptions

Apart from action and state descriptions, another frequent aspect in the human generated instructions are *referential descriptions* of entities followed by anaphorical reference to avoid redundancy. They are extended over many sentences or subclauses and refer to various antecedents. These kinds of instructions that include a set of references, are structured in a very complex way, so that reference resolution might be quite difficult for robotic teammates. The example based on Task C in Figure 1

(e)"How close can you get to the victim and if is it safe to get to him, to assist him and how badly is he hurt"

shows an instruction that is split into many subclauses. The anaphors *him* and *he* are extended over the whole instruction and refer to one antecedent to *victim*. In the results the referential descriptions have mostly been used to refer to the victim given in Task C. A small amount of those was used to refer to inanimate entities in the world. In Figure 2 the expressions used in the results by the participants based on the different landmarks is illustrated.

A reason for the small amount of references to describe inanimate entities might be that humans are used to utter expressions such as "there" to refer to regions or areas. In the results it was quite observable that those kinds of expressions have been used by 48 subjects to describe regions or areas very vague without indicating entities explicitly. An example on Task C is given with

(g) "Go to the two trees at the top of the mountains. Are there victims at your present location?"

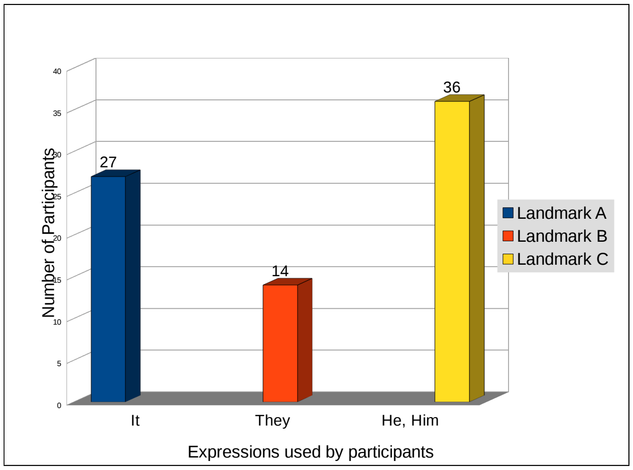

Fig. 2. Referential descriptions applied in the contributions by participants

TABLE I. FREQUENCY OF SPECIAL CHARACTERS BASED ON THE GIVEN LANDMARKS

Landmark	Comments	Num of Participants
A	huge tree, largest tree, big tree,tall pointy tree	40
B	large rocks, big boulders, pile of rocks	36
C	across snow pack to peak, all trees at the top of the hill, valley between the mountains, up to mountain two trees	37

D. Different Viewpoints

Another essential aspect highlighted in the results, are the use of *various viewpoints* applied by the participants to their instructions. Around 35 participants instruct their team by using spatial directions based on their teams position, e.g. *Move to your right*. The reason might be when humans work alongside humans they are on a par with their teammates while working with robotic teammember is more subordinated. The task execution is more delegating than collaborating.

By investigating these results it was quite clear that the hypothesis *H1* is satisfied by around 36-40 participants. Humans are tend to use specific visual descriptions for entities when instructing or guiding their teammates. In Table I some of the most used expressions (based on the landmarks in Figure 1) applied by the participants in their utterances is given.

In order to describe the location and the surroundings of the several landmarks the participants uttered expressions that indicated the sizes and shapes of entities and the relation of dependencies between entities. An example is given by *all trees at the top of the hill* that specifies the location of the trees at the top of another location.
While checking hypothesis *H2*, it showed that most of the participants did not consider the conditions of the terrain in their instructions, e.g. obstacles along their way. Around 22 subjects included the surroundings in their utterances with expressions such as *" Be careful"*, *"Be aware"* or *"Be cautious"*. The reason might be that humans are expecting from other humans cognitive capabilities such as awareness or perception when performing tasks. Something that is *naturally* given by humans and does not have to be explicitly mentioned. As long as the obstacle is not hindering the task execution, humans do not see this information essential to share. This aspect is very important when performing tasks indoor as well as outdoor. Robotic teams need to have these kind of capabilities in order to be able to replan or reroute themselves when an obstacle is along their path. So, the hypothesis *H2* is also satisfied based on the given tasks and their results.

The results show that the following issues have to be considered to enable a smooth interaction between humans and robots:

- Reference Resolution: Robots must be able to recognize dependencies in the sentences and resolve the references that are extended over the whole instruction.
- Sequences of Action and State Descriptions: Robots need to consider and interpret instructions with various length and with pre- as well as postconditions.
- Distinctive Descriptions: Robots need knowledge about distinctive descriptions given in the world such as "eyecatching" entities.
- Understanding Relations and the Semantics between Entities: Robots need to understand and interpret qualitative spatial expressions in instructions based on the given conditions of the world.
- Taking Perspective: Robots have to consider the different viewpoints when interpreting instructions given by a human. Humans are often used to refer to objects that are in their visibility range or in front of them.

The results of our study showed that humans instruct other humans by using expressions in a way that are unknown for robotic systems. They need special skills to comprehend and process those expressions. When interpreting referencial expressions in instructions into the task context, for example, robots must be able to recognize those dependencies in the sentences and resolve the references, accordingly.
In order to be able to understand and interpret instructions at a human level, we have extended our current system with the results of this section to improve it. This system will be presented in the next section.

V. EXPERIMENTAL EVALUATION

In this section we will show how the results of the study can improve the capabilities of systems during task interpretation. By extending our task interpretation system, we will examplarily show how our system is able to handle ambiguity and vagueness in human instructions without requesting additional information from humans.

A. Interpretation of High-level Instructions through Task Interpretation System

Using the *VOD-structure* allows us to easily define rules to translate high-level instructions into action descriptions comprehensible for robotic systems. Natural language instructions such as

(a) Go across the snow pack to the peak

Algorithm 1 Parametrized Action Description (1)

```
(an action
  (:type move)
  (:viewpoint team-leader)
  (:direction across)
  (a region
    (visible
      (a region
        (:name snow-pack)
        (:to
          (a region
            (:name peak)))))))
```

can be directly interpreted as a high-level specification and can be easily translated into an action description (see Algorithm 1) executable by robotic systems.

This description uses essential components of the Cognitive Robot Abstract Machine (CRAM) [1], [10], called *designators* that describe various parameters and offer several forms of implementations for objects, locations and actions. In the listing above, the interpreted command is translated into an action designator that includes three location designators, equated with regions. All of them contain a set of symbolic constraints that are conditions to restrict the search space. These constraints are presented by colons followed by expressions, such as *:type*, *:viewpoint*, *:direction*, *:name* and *:to*. This description was extended with the *:viewpoint* tag in order to consider the certain viewpoint during the action execution.

In order to be able to directly interpret instructions given by the *VOD-structure* into action descriptions, we have extended our existing task interpretation system introduced in [17] with a new component, called Human-Machine Interface (HMI). This component consists of a parser and an interpreter that includes algorithms for natural language processing and interpretation developed for this use case.
The parser that we use is called temporal logic and dynamic logic NL parser (TLDL) [2]. It takes lexical items and maps them to temporal and dynamic logic expressions that represents goals and actions specified in the NL directive, respectively. For the purposes of the current search task, we have extended this parser based on our results to process dependencies such as in spatial relation descriptions. To handle those depedencies given with

(h) Go right of the tree next to the rock and take a picture.

we extended the parser by adding a *sequential-connector* to mark that parts of the instruction belongs together. In the instruction above the connector is added between the pronouns "tree" and "next" to indicate that both parts of the instruction belong to the same sentence. After parsing this instruction, the system is generating formulas such as

move(right, tree)<=(next-to,rock);*take(picture)*.

The action *take(picture)* is interpreted as a definition of visibility that includes the meaning of detecting something. By adding the connector which is presented as "<=" in the formulas instructions based on various length of sequences can be interpreted. After the parser mapped the natural language command into logical expressions, the results are forwarded to the interpreter. The interpreter is equipped with algorithms to reason about spatial relations, dependencies between sentences and reason about entities represented in the different viewpoints. In order to fill the knowledge gaps and resolve ambiguities in the knowledge, the HMI component retrieve and access information of the knowledge processing system (KnowRob) [13]. KnowRob enables the robotic agent to retrieve background information about entities within the world and reason about the execution to generate more flexible and robust behavior. By retrieving information, the HMI is able to ground the expressions in the instructions to the perception of the robotic systems and resolve the complex linguistic and visual descriptions as well as create semantically annotated descriptions.

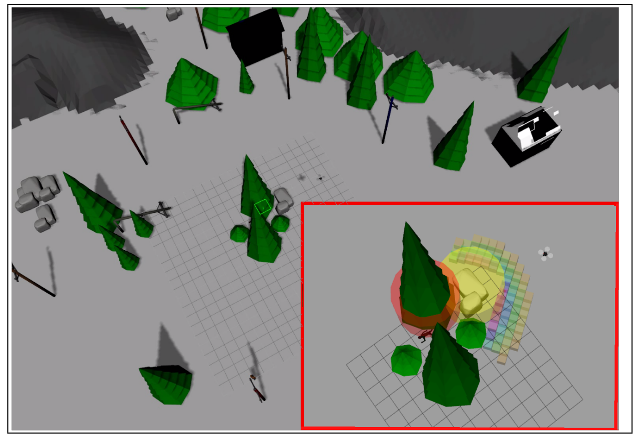

Fig. 3. Task interpretations in a simulated world with the agents, e.g. human colored yellow and a quadrotor, and the visualization of sampling-based goal positions for "Go right of the tree next to the rock". Mechanism uses the rock as reference point and calculates robot goal positions for the closest tree that has on his right side a rock.

To translate the abstract and qualitative action descriptions into numeric action parameters which the robot's navigation subsystem can use, our system introduced in [17] generates a distribution of goal positions that satisfy the qualitative description. In Figure 3 the results of this approach depicting a simulation environment including a human and robot team based on the qualitative description "right of a tree next to a rock" is shown. After the interpretation of the instruction, we create a sampling-based map of possible goal positions where the robot can fly in order to see the victim.

Note that the system generates positions on the right side of the tree because one constraint is to calculate positions "next to the rock" and "right of the tree" from the viewpoint of the human. The information related to the position of the tree and the rock are obtained from the world model. After the sampling-based map is generated out of all samples one will be picked out and sent to the robot in Gazebo in order to fly to this position.
Additionally, we have also developed inference mechanisms to interpret spatial dependencies based on their appearance in the instruction, such as in *(h)*. Those dependencies have to be correctly interpreted as follows: go next to a rock that is right of a tree. In order to interpret this instruction correctly, we integrated our inference mechanisms into the HMI component to ensure a correct interpretation as follows: After the team leader instructs the team we are generating

inferences on objects labeled with rocks that are located close to the instructor based on his field of view. Rocks behind him are not considered by our mechanisms because when we are giving instructions we just consider those objects that are in our field of view and not behind us. After the inferences are generated, our mechanisms will validate each inference if it satisfies the condition “a rock right of a tree”. If a match is found, it will be selected and sampling-based positions will be calculated. In order to understand how the different algorithms and the inference mechanisms work, we generated a short video that shows the functionality of our system and can be seen here [1].

All these extensions are essential to ensure natural taskability between human and robot teammates. On the one hand, these extensions enabled humans to intuitively instruct their teammates without considering which visual and semantical descriptions they have applied on entities in the environment, e.g. tree or lake. On the other hand, robots were able to understand the intention and behavior of humans and perform their instructions into the context of the task.

VI. Conclusion and Future Work

In this paper we introduced an experimental setup of an empirical study in a visual search and rescue task application for mixed human-robot teams and presented the empirical results. We introduced the results of our study and integrate them into our task interpretation system that was extended with a Human-Robot Interface to translate natural language instructions into action plans understandable and executable by robotic team members. It is obvious that robots are not able to make relations or to understand the semantics of sentences due to their lack of language skills.

By this proposed approach we are showing that natural language systems needs to be equipped with specific abilities that helps robots to interpret instructions into the context of the task in order to perform actions correctly. It is important to share a common ground based on understanding language at human-level in order to support robots in understanding human instructions. The results of this work can be used as guidelines for dialogue systems in order to improve and optimize systems as well as to help robots to get a better understanding of human intention and behavior while interpreting instructions.

The current system is, of course, only a start and more work on improving the human and robot dialogue is required to handle referential expressions of entities and cases of ambiguities. In the future we will extend our system with NL processing components that can determine and resolve more complex linguistic and visual descriptions and that can resolve ambiguities in the instructions by ask questions to clarify goals or by querying different knowledge bases.

References

[1] M. Beetz, L. Mösenlechner, M. Tenorth, and T. Rühr. Cram – a cognitive robot abstract machine. In *5th International Conference on Cognitive Systems (CogSys 2012)*, 2012.

[2] J. Dzifcak, M. Scheutz, C. Baral, and P. Schermerhorn. What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA '09)*, Kobe, Japan, May 2009.

[1] https://www.dropbox.com/s/akir91p2tgd0n5z/conf.mpg?dl=0

[3] T. W. Fong, J. Scholtz, J. Shah, L. Flueckiger, C. Kunz, D. Lees, J. Schreiner, M. Siegel, L. Hiatt, I. Nourbakhsh, R. Simmons, R. Ambrose, R. Burridge, B. Antonishek, M. Bugajska, A. Schultz, and J. G. Trafton. A preliminary study of peer-to-peer human-robot interaction. In *International Conference on Systems, Man, and Cybernetics*. IEEE, October 2006.

[4] A. Green, H. Hüttenrauch, M. Norman, L. Oestreicher, and K. S. Eklundh. User centered design for intelligent service robots. 2000.

[5] G. Hoffman and C. Breazeal. Collaboration in human-robot teams. In *Proceedings of the AIAA 1st Intelligent Systems Technical Conference*, 2004.

[6] E. Krause, M. Zillich, T. Williams, and M. Scheutz. Learning to recognize novel objects in one shot through human-robot interactions in natural language dialogues. In *Proceedings of Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[7] G. J. Kruijff, M. Janicek, S. Keshavdas, B. Larochelle, H. Zender, N. Smets, T. Mioch, M. Neerincx, J. Van Diggelen, F. Colas, M. Liu, F. Pomerleau, R. Siegwart, V. Hlavac, T. Svoboda, T. Petrıcek, M. Reinstein, K. Zimmermann, F. Pirri, M. Gianni, P. Papadakis, A. Sinha, B. Balmer, N. Tomatis, R. Worst, T. Linder, H. Surmann, V. Tretyakov, S. Corrao, S. Pratzler-Wanczura, and M. Sulk. Experience in system design for human-robot teaming in urban search and rescue. In *Proceedings of 8th International Conference on Field and Service Robotics*, STAR. Spring Verlag, 2012.

[8] T. Laengle, T. Hoeniger, U. Rembold, and H. Woern. Cooperation in human-robot-teams. *Proceedings of the ICI and C'97, Volume 3 of 3*, pages 1306–1314, 1997.

[9] L. Marconi, C. Melchiorri, M. Beetz, D. Pangercic, R. Siegwart, S. Leutenegger, R. Carloni, S. Stramigioli, H. Bruyninckx, P. Doherty, A. Kleiner, V. Lippiello, A. Finzi, B. Siciliano, A. Sala, and N. Tomatis. The sherpa project: smart collaboration between humans and ground-aerial robots for improving rescuing activities in alpine environments. In *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, College Station, Texas, USA, Nov. 5-8 2012.

[10] L. Mösenlechner and M. Beetz. Parameterizing actions to have the appropriate effects. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, USA, September 25-30 2011.

[11] I. R. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto. An affective mobile robot educator with a full-time job. *Artificial Intelligence*, 114(1):95 – 124, 1999.

[12] N. Roy, G. Baltus, D. Fox, F. Gemperle, J. Goetz, T. Hirsch, D. Margaritis, M. Montemerlo, J. Pineau, J. Schulte, and S. Thrun. Towards personal service robots for the elderly. In *Workshop on Interactive Robots and Entertainment (WIRE 2000*, 2000.

[13] M. Tenorth and M. Beetz. KnowRob – A Knowledge Processing Infrastructure for Cognition-enabled Robots. Part 1: The KnowRob System. *International Journal of Robotics Research (IJRR)*, 2013. Accepted for publication.

[14] K. M. Tsui, M. Desai, H. A. Yanco, and C. Uhlik. Exploring use cases for telepresence robots. In A. Billard, P. H. K. Jr., J. A. Adams, and J. G. Trafton, editors, *Proceedings of the 6th International Conference on Human Robot Interaction, HRI 2011, Lausanne, Switzerland, March 6-9, 2011*, pages 11–18. ACM, 2011.

[15] T. Williams and M. Scheutz. Power: A domain-independent algorithm for probabilistic, open-world entity resolution. In *Proceedings of IROS*, 2015.

[16] M. Xin and E. Sharlin. Exploring human-robot interaction through telepresence board games. In Z. Pan, A. D. Cheok, M. Haller, R. W. H. Lau, H. Saito, and R. Liang, editors, *Advances in Artificial Reality and Tele-Existence, 16th International Conference on Artificial Reality and Telexistence, ICAT 2006, Hangzhou, China, November 29 - December 1, 2006, Proceedings*, volume 4282 of *Lecture Notes in Computer Science*, pages 249–261. Springer, 2006.

[17] F. Yazdani, B. Brieber, and M. Beetz. Cognition-enabled robot control for mixed human-robot rescue teams. In *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13)*, 2014.