

# AN NLP-BASED APPROACH FOR IMPROVING HUMAN-ROBOT INTERACTION

Yilmaz Kilicaslan<sup>1</sup> and Gurkan Tuna<sup>2</sup>

<sup>1</sup>*Department of Computer Engineering,  
Faculty of Engineering, Trakya University,  
Edirne, Turkey*

<sup>2</sup>*Department of Computer Programming,  
Edirne Vocational School of Technical Sciences, Trakya University,  
Edirne, Turkey*

## Abstract

This study aims to explore the possibility of improving human-robot interaction (HRI) by exploiting natural language resources and using natural language processing (NLP) methods. The theoretical basis of the study rests on the claim that effective and efficient human robot interaction requires linguistic and ontological agreement. A further claim is that the required ontology is implicitly present in the lexical and grammatical structure of natural language. The paper offers some NLP techniques to uncover (fragments of) the ontology hidden in natural language and to generate semantic representations of natural language sentences using that ontology. The paper also presents the implementation details of an NLP module capable of parsing English and Turkish along with an overview of the architecture of a robotic interface that makes use of this module for expressing the spatial motions of objects observed by a robot.

## 1 Introduction

Autonomous mobile robots can come into physical proximity with other autonomous robots, people, and objects and can negotiate in a dynamic environment. At the same time, having learnt about themselves and their world they make decisions to take actions [1]. Since autonomous mobile robots operate in the physical world, they must adjust their decisions sensibly and safely to account for their abilities and the challenges in the environment [2]. Therefore, the robots built for interacting with other robots and people need to be designed accordingly.

This paper argues that effective HRI can be achieved through natural language communication and presents the details of an HRI system developed to this effect. Efficient communication requires a certain degree of agreement at two levels: the level

of language and that of ontology. First of all, the participants of a communication act must use either the same language or languages that are translatable to each other. Secondly, even though the participants use the same language, they also have to agree about their perspective of reality. If we are not solipsist in our philosophical stance towards reality, we should somehow admit that we live in the same reality. However, this does not mean that we perceive and conceive this reality in the same way. Such differences may simply arise out of differing perceptions and/or conceptualizations of the same piece of reality. Gruber refers to ‘a specification of a conceptualization’ as an ‘ontology’ [3]. Hence, in Gruber’s terms, efficient communication requires ontological agreement. In fact, ontological engineers approach the matter in Gruberian spirit in a similar way, as the following quote reflects from the

website owlseek.com reflects:

*We can never know reality in its purest form; we can only interpret it through our senses and experiences. Therefore, everyone has their own perspective of reality. An ontology is a formal specification of a perspective. If two people agree to use the same ontology when communicating, then there should be no ambiguity in the communication.*

A further claim of the paper is that an ontology inheres in the lexical and grammatical structure of natural language. The paper offers some NLP techniques to uncover (fragments of) the ontology hidden in natural language and to generate semantic representations of natural language sentences using that ontology. The paper also presents an NLP module along with a robotic interface that makes use of this module for expressing the spatial motions of objects observed by a robot.

Following this section, the paper is organized as follows. Section 2 presents a review of the current trends in human robot interaction. Section 3 explains the design and implementation details of the NLP module developed for effective HRI. Section 4 is a discussion of how to derive an ontology from natural language. Section 5 offers an overview of the architecture of the proposed NLP-based robotic interface. Open research issues are discussed in Section 6. Section 7 concludes the paper.

## 2 Related Work

HCI has been around for a long time. On the other hand, HRI is a much recent topic. Brown *et al.* in [4] propose a cognitive robot to study the resolution of ambiguity in HRI. Yanco and Drury in [5] propose a taxonomy for classifying HRI. According to [5], the categories of the taxonomy are task type, task criticality, robot morphology, ratio of people to robots, composition of robot teams, the level of shared interaction among teams, interaction role, type of human-robot physical proximity, decision support for operators, time/space taxonomy, and autonomy level.

Most HRI applications have been focused on teleoperation, service robots, robot hosting, and physical HRI. Yanco *et al.* in [6] describe robot systems designed for urban search and rescue (USAR) and report the results of usability tests conducted

to compare the interfaces developed for HRI. In the work in [7], Casper and Murphy report an analysis which was performed on the data collected during the The World Trade Center (WTC) rescue response. In [8], the details of COGNIRON project, a European Union project to develop a social robot companion, are given. Tellex and Roy in [9] propose a speech-controlled wheelchair which understands high-level natural language commands. A robot penguin with engagement capabilities is introduced in [10]. In [11] Khatib *et al.* describe physical HRI which involves physical (such as haptic, force and neural) interactions between humans and robots.

Semantic knowledge in robotics can be characterized by the need for an explicit representation of knowledge inside the robot and the need for grounding the symbols used in this representation in real physical objects, parameters, and events [12]. Since semantic knowledge reduces the possibility to reuse knowledge, to reason on it, or to share it with robots, devices, or humans, it is found in many socially-intelligent robots [13]. It not only improves the performances of autonomous robots, but is also an effective approach in HRI, especially when the presence of humans is fundamental. It can be applied to several applications. A natural language interface with a knowledge base containing spatial information to describe static situations and actions for mobile robots is proposed in [14]. The use of spatial relations to ease human-robot communication is investigated in [15]. Kruijff *et al.* in [16] propose a multi-layered conceptual spatial mapping to provide a common ground for human-robot dialogues. In [17] Hasanuzzaman *et al.* define a knowledge-based management system for person-centric gesture interpretation. An unsupervised learning method used to identify and build an object ontology without using explicit symbolic representations is proposed in [18].

Our study differs from the related semantically oriented approaches in that it focuses on the ontology inherent in natural language and places special emphasis on the use of NLP techniques to uncover that ontology and thereby to semantically process natural language sentences.

### 3 An NLP Module for Parsing English and Turkish

Part of effectiveness of HRI rests on the use of natural languages as a communication means. This will at least remove from human operators the burden of thinking and acting within the limitations of an artificial language. The HRI system is equipped with a natural language processing (NLP) module that is capable of parsing English and Turkish. Statistical and rule-based approaches to parsing are two main strategies employed in NLP. The HRI system has recourse to the C&C tools to parse inputs from English. The C&C tools contain an NLP system capable of extracting a grammar of a language from a corpus using statistical learning algorithms and certain syntactic and semantic formalisms [19]. It is a fact that the statistical modelling of language has led to high levels of robustness and efficiency in NLP systems and thereby made large-scale language processing a possibility [20], [21]. However, as a sufficiently large and appropriately annotated corpus is not available for Turkish, the HRI system handles inputs from this language using rule-based mechanisms.

The NLP system coming with the C&C tools is built around a Combinatorial Categorical Grammar (CCG) parser [22]. The grammar of English used by the parser is extracted from CCGbank, a CCG version of the Penn Treebank [23]. Another component of the C&C tools, namely Boxer, is responsible for the semantic interpretation of English. Boxer takes CCG derivations output by the C&C parser and generates semantic representations as box-like structures of Discourse Representation Theory, DRT [24], which are known as Discourse Representation Structures (DRSs). A DRS has two critical components:

- a set of *discourse referents* representing entities which are under discussion; and
- a set of *DRS conditions* representing information about discourse referents.

Default DRS outputs are in Prolog format, with discourse referents represented as Prolog variables and DRS conditions as Prolog's compound terms. Fig. 1 shows a DRS output produced by Boxer for the sentence *But Mr. Barnum called that a worst-case scenario*:

```

| x0 x1 x2 x3
|-----|
| named(x0,barnum,per)
| named(x0,mr,t11)
| thing(x1)
| worst-case(x2)
| scenario(x2)
| call(x3)
| but(x3)
| event(x3)
| agent(x3,x0)
| patient(x3,x1)
| theme(x3,x2)
|-----|

```

Figure 1. A DRS output by Boxer

DRSs can be thought of as partial representations of the speaker and hearer's (shared) interpretation of a discourse as it unfolds over time. They are partial because, as cognitive linguists emphasize, the interpretation of an utterance is not fully determined by its linguistic form:

*A cognitive approach to meaning construction holds that the interpretation of language is integrative, elaborative and inherently conceptual in nature. On this view, interpretation, which is to say meaning construction, is not simply the result of compositionally adding linguistic items. Rather utterances- lexical items and the syntactic configurations in which they occur - provide only minimal prompts for meaning construction. Language vastly underdetermines the rich interpretations normally assigned to even simple, de-contextualized sentences; sentential interpretation results from the integration and elaboration of these minimal linguistic cues at the conceptual level [25].*

### 4 Uncovering Fragments of Ontology Inherent

The ontologies that the discourse participants use are crucial sources of conceptual content of utterances that they interpret and the extent to which they understand and/or agree with each other crucially depends on the agreement of their ontologies. The question to ask now is how to get to an ontology for an HRI system. Our answer is that any natural language can be the place where an ontology can be uncovered.

As it is pointed out in [26], “[t]he relation between ontologies and language is currently at the forefront of natural language processing (NLP)” and “[o]ntologies, as widely used models in semantic technologies, have much in common with the lexicon.” (p.i) In fact, formalizing the lexicon of a natural language will result in a general-purpose ontology, more or less shared by the speakers of that language. Kilicaslan and Guner in [27] propose to develop such an ontology using Wordnet, which can be considered to be a lexicon of the English language. They have developed a concept lattice generator that automatically transforms Wordnet to an ontology. Below is a lattice-based fragment of that ontology:

Kilicaslan and Guner formalize their ontology using Formal Concept Analysis (FCA). FCA is a mathematical theory of concepts and concept hierarchies [28], where a concept is a pair consisting of a set of objects, which is the ‘extent’, and a set of attributes, which is the ‘intent’, such that:

- the extent consists of all objects that share the given attributes and
- the intent consists of all attributes shared by the given objects.

A formal ontology derived by FCA is a concept hierarchy where the set of all concepts is ordered by a subconcept-superconcept relation, which is a particular order relation denoted by  $\leq$ . If  $(O1, A1)$  and  $(O2, A2)$  are concepts, the former is said to be a subconcept of the latter, i.e.  $(O1, A1) \leq (O2, A2)$  if:

$$O1 \subseteq O2 \Leftrightarrow A1 \supseteq A2.$$

A set ordered in this way is called a concept lattice. A concept lattice can be drawn as a diagram in which concepts are represented by nodes interconnected by lines going down from superconcepts to subconcepts, as shown in Fig. 3.

Even though ontology in philosophy may presuppose some kind of uniqueness, as philosophical ontology should cover all of the world as the theory of existence, the term ‘ontology’ is used in a more limited and practical sense in computer science. As Takeda and Nishida point out in [29], various types of ontologies can be constructed depending on the application in which they are used.

There are two types of ontologies employed by our HRI system: domain-specific ontologies and thematic ontologies. For each application domain, a specific ontology is developed in FCA so that the objects and attributes peculiar to the application can appear in the shared ontology. For example, the following concept lattice will be derived for a context where the objects are John (j), Fido (f) and Tweety (t) and the attributes are animate, smart, two-legged and furry, as shown in Fig. 4.

Thematic relations are semantic correlates of grammatical relations encoded by case morphology or syntactic position. When scrutinizing the content side of this correlation, our point of departure has been the theory of semantic structures developed by Jackendoff in [30]. In his theory:

*meaning*, like phonological structure, is organized into independent but interacting tiers, each of which contributes a different class of conceptual distinctions to meaning as a whole. (p. 2)

Thematic (conceptual) roles are the main building blocks of these tiers. This study is concerned with two of the tiers: the *thematic tier* dealing with motion and location and the *action tier* dealing with Actor-Patient relations. The focus of the study is on the thematic tier. Therefore, only that tier will be analysed and the action tier will be left unanalysed in what follows.

Our re-formalization of Jackendoff’s theory is based on FCA. Both tiers are structured as concept lattices where the attributes are certain relevant thematic roles and the objects are entities playing these roles in a given situation. Below is the set of the thematic roles that constitute the thematic tier:

{Location, Source, Goal, Path}

We argue that a thematic tier has a mathematical structure that is richer than that of a flat set of thematic roles. A *flat set* is a set that contains atoms (i.e. non-sets) as elements. Thus, the set above, which we call the locative roles, should be considered only as an abstraction over the structure of the tier. If we look at the definitions of the thematic roles in the set, we can see that they are organized in a lattice hierarchy.

**Location:** the position where an entity is.

**Source:** The position from which an entity moves.



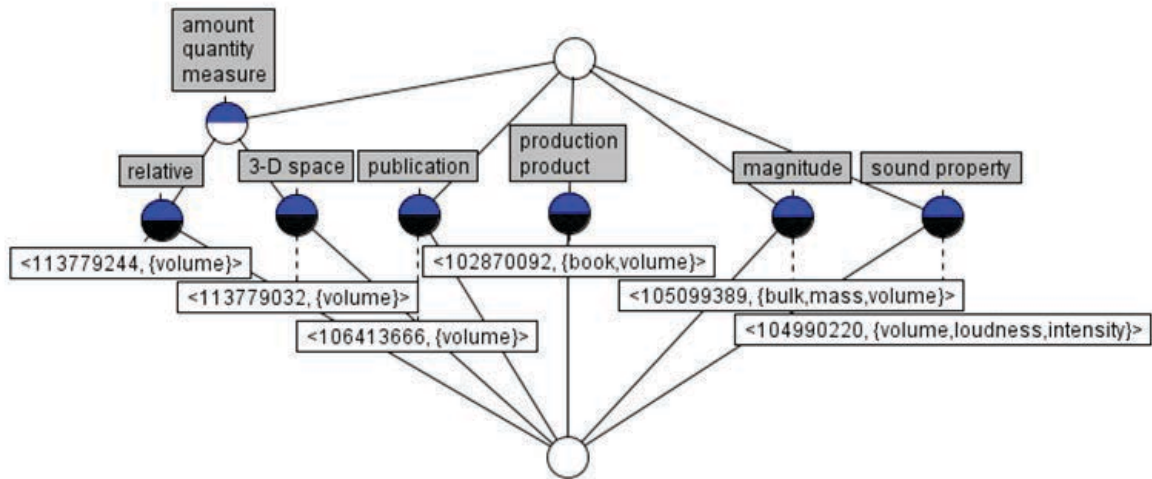


Figure 2. A fragment of a lattice-based ontology derived from Wordnet

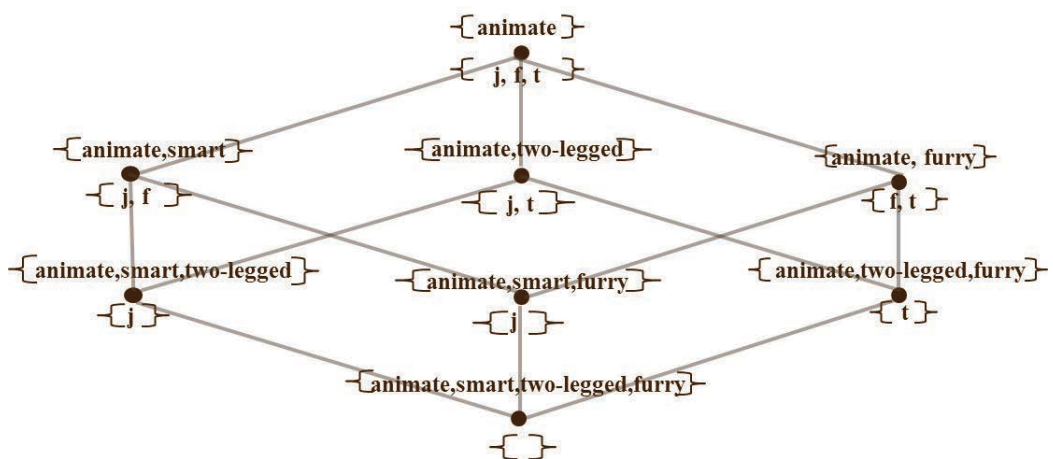


Figure 3. A domain-specific ontology

**Goal:**The position to which an entity moves.

**Path:**The trajectory along which an entity moves.

Consider now a situation where a child walked from a living room to a kitchen through a hallway in a house. The house, the living room, the kitchen and the hallway played respectively the roles Location, Source, Goal and Path in this situation. But, the living room, the hallway and the kitchen need also be considered as bearing the thematic role Location, as each was a place where the child was at a different time interval of the situation. Moreover, the hallway functioned as the goal to which the child moved at the beginning of the situation and as the source from which the child came out at the end. All these facts can be captured by means of the lattice in Fig. 4, where the house, the living room, the kitchen and the hallway are referred to as *h*, *r*, *k*, and *w*, respectively.

In some cases an object might bear both a locative role and a role of the action tier. Consider a situation where a girl hits a boy with a ball. The ball bears the role Theme. It moves from the girl to the boy. In this respect, the girl and the boy play the roles Source and Goal, respectively. However, the girl and the boy are also role-players in the thematic tier: they are Actor and Patient, respectively. Ignoring the internal structure of the action tier, we can represent such multi-role cases in Fig. 5 (where *g* and *b* denote the girl and the boy, respectively). This example illustrates a case where the thematic and action tiers interact, with an object coming into contact with another after moving through space.

## 5 An Overview of the Robotic Interface

An overview of the proposed semantic-driven system architecture is shown in Fig. 6. In the proposed system, visual feedback provides inputs to the system. Then, visual feedback is processed through a semantic-driven subsystem complemented by the NLP module explained in Section 3. Finally, commands are sent to a task management subsystem through a dialogue screen and a robot controller performs appropriate steps such as driving the robot forward, stopping the robot, and moving its arm. The overall flow of the visual feedback subsystem is shown in Fig. 7 and the flow of

the object tracking function is shown in Fig. 8. In the visual feedback subsystem we focused on the object tracking function since object tracking is the main part of most visually driven systems.

The abovementioned steps are performed to track an object in the application. The implementation of the proposed architecture has been realized using the Robot Operating System (ROS) [31] on Ubuntu and deployed on the prototype robot shown in Fig. 9. Fig. 10 shows an image obtained by the robot's camera. Whenever an object which is in front of the robot's camera moves, the application writes a appropriate message on the console. Yet, we need to integrate the console of dialog management subsystem into the ROS-based robot controller shown in Fig. 10. They are currently independent applications except for the message handling functions supported by ROS.

## 6 Open Research Directions

Current version of the application uses reference points in video frames based on local coordinate axes so when the application is deployed to real robots, it may provide inaccurate results. Therefore, the positions of objects in a video frame can be processed by comparing with the positions of the same objects in the previous video frame.

In multi-robot systems, the decision given by an initiating robot can be improved by the decision of other robots. For instance, due to the lack explained above, when robot A says "robot B moved up or moved to  $(x_2, y_2)$  from  $(x_1, y_1)$ ", robot B can inform robot A that it did not move at all. Therefore, a multi-master decision making schema can be used to improve the decision making process described in this study.

Due to the implementation related constraints and the design considerations, the following design issues were left as future research directions:

- a statistical evaluation to indicate a 95% confidence level for the given decisions,
- measuring how well the multi-robot system learns the tasks assigned to it by human operators,
- varying the parameters in the system to measure

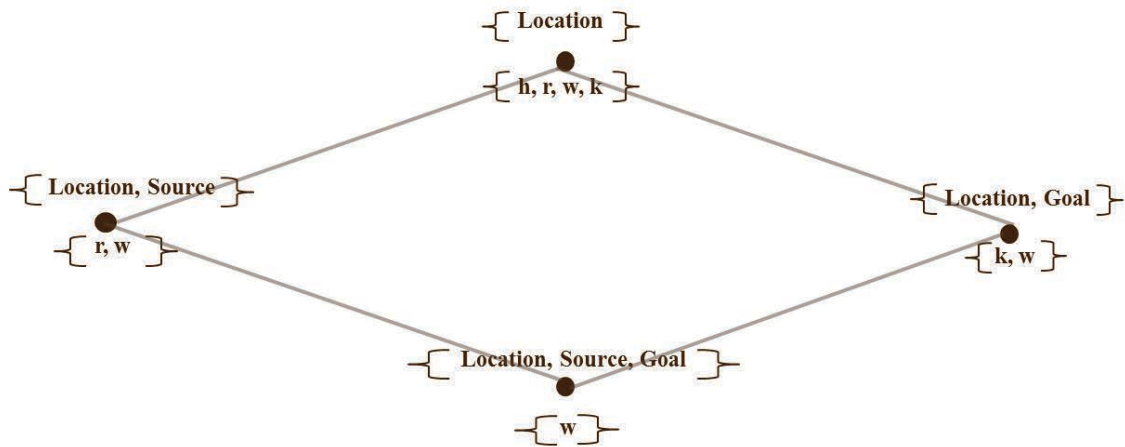


Figure 4. An example of the thematic tier

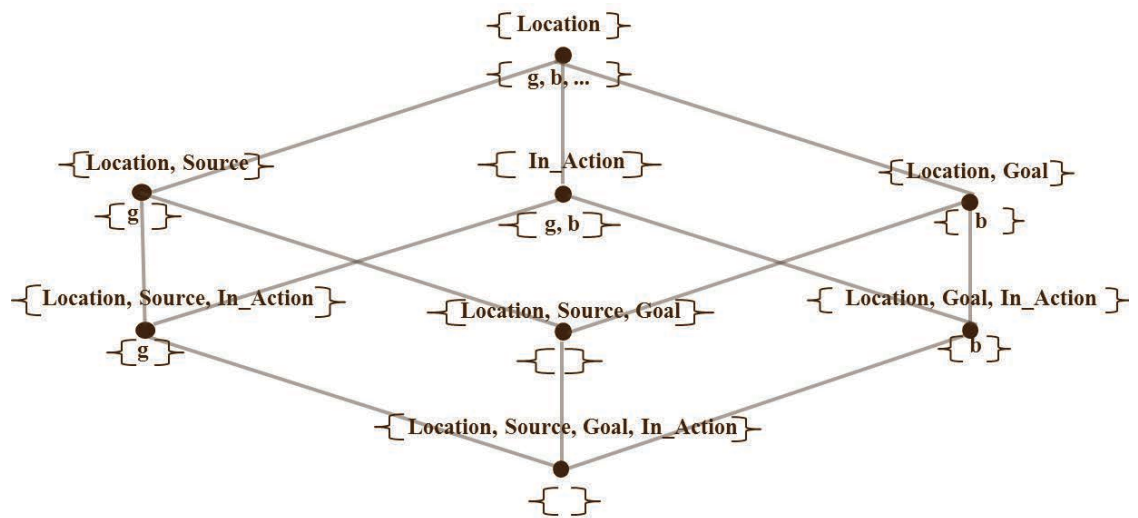
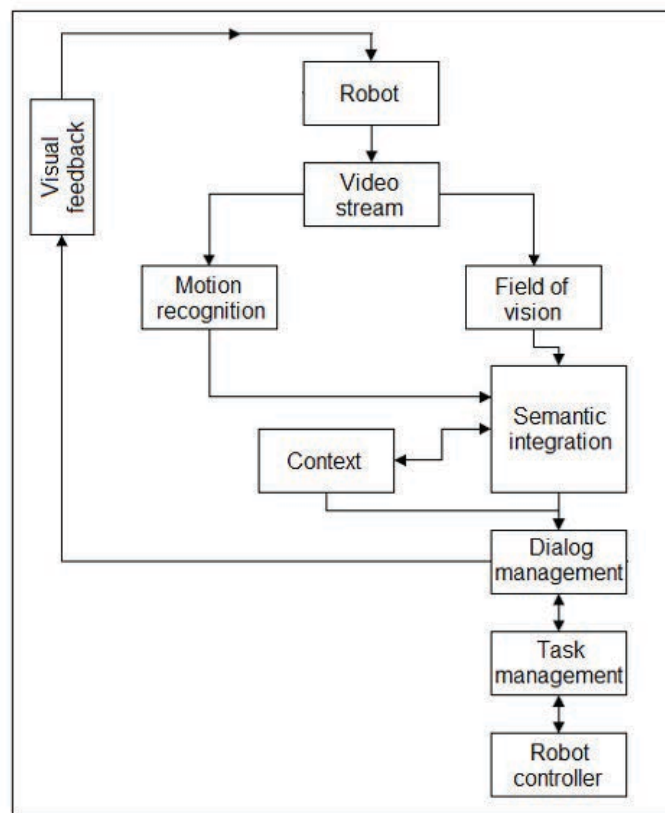


Figure 5. The point of connection between the two tiers.



**Figure 6.** An overview of the system architecture



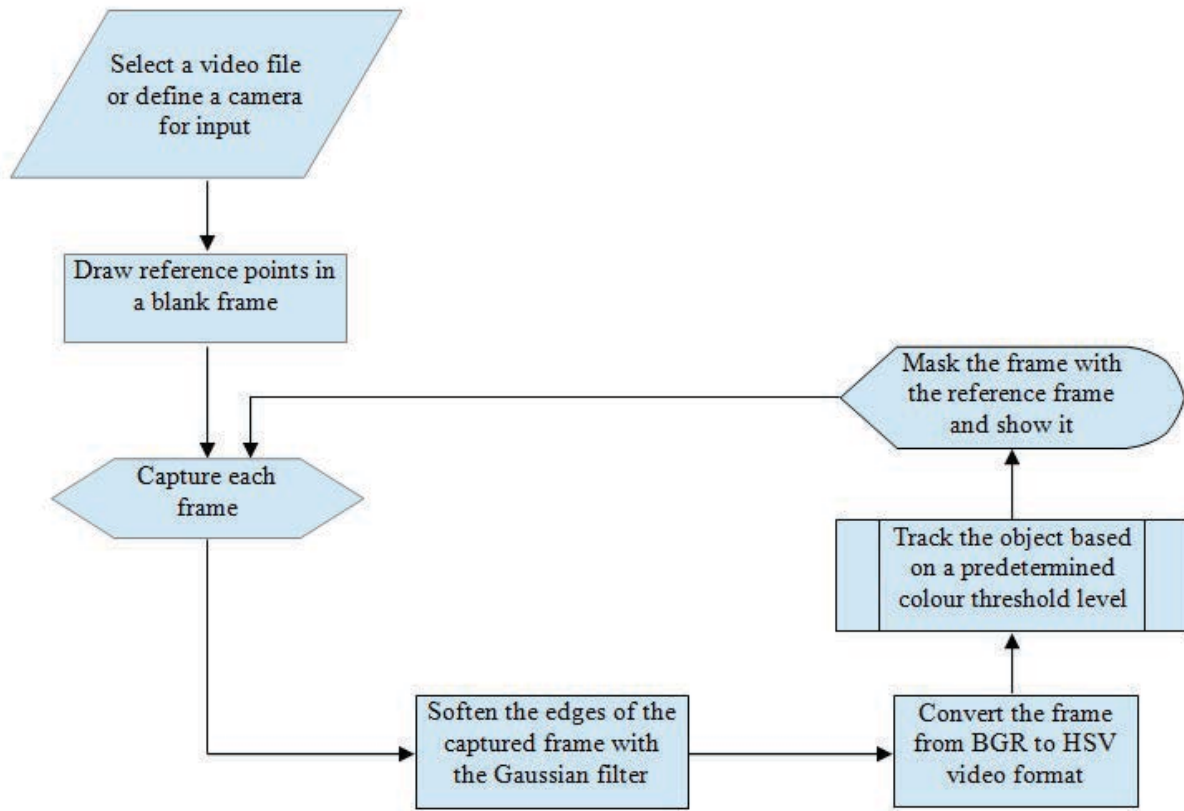


Figure 7. Overall flow of the application

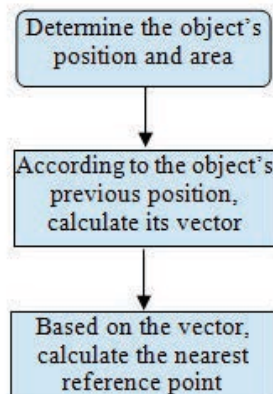


Figure 8. Flow of the object tracking function

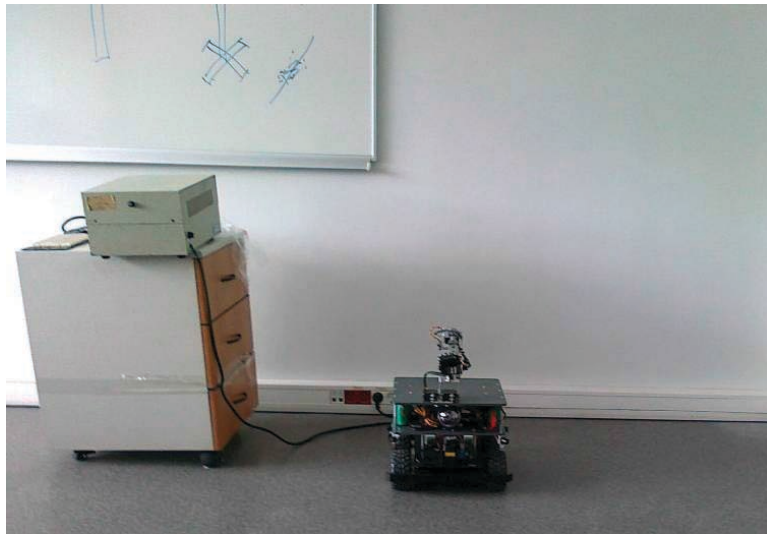


Figure 9. The prototype robot

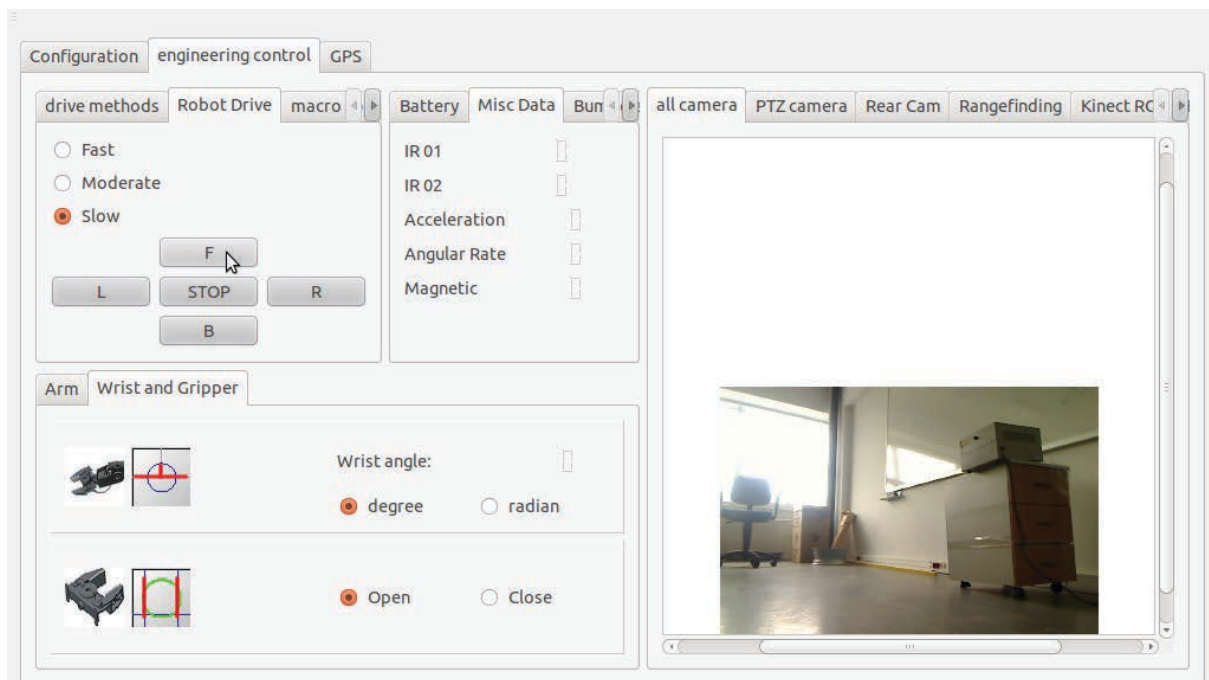


Figure 10. An image obtained by the robot's camera

the impact of the parameters on the accuracy of the system,

- introducing vision system recognition errors to simulate real world conditions,
- introducing audio system recognition errors and background noises to measure the efficiency of the system in a real world environment.

## 7 Conclusion

Efficient communication is the key requirement of coordination and task allocation strategies in multi-robot systems. When using technologies to enable communication in multi-robot systems, data must be represented in an appropriate and common form.

This study argues that linguistic and ontological agreement is necessary to improve the HRI abilities of visually-driven robots. A vision-based architecture backed by an NLP-based subsystem improves both the HRI abilities of a robot and coordination efficiency of a multi-robot system. Both the linguistic and ontological capabilities of such systems can be acquired using NLP techniques.

## 8 Acknowledgments

This study was partly funded by the R&D, Innovation and Industrial Application Support Programme of Small and Medium Enterprises Development Organization in Turkey with the research project entitled "Otistik ve Zihinsel Engelli Bireyler İçin Dogal Dil İfadelerinin Bilgisayar Ortamında Gorselleştirilmesi" (Computer-based Visualization of Natural Language Expressions for Individuals with Autism and Mental Retardation).

## References

- [1] J. Cheng, W. Bian, and D. Tao, Locally regularized sliced inverse regression based 3D hand gesture recognition on a dance robot, *Information Sciences*, vol. 221, 2013, pp. 274-283.
- [2] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, A Model of Proximity Control for Information-Presenting Robots, *IEEE Transactions on Robotics*, vol. 26, no. 1, 2010, pp. 187-195.
- [3] T. R. Gruber, A Translation Approach to Portable Ontologies, *Knowledge Acquisition*, vol. 5, 1993, pp. 199-220.
- [4] F. Brown, A. Agah, J. Gauch, T. Schreiber, and S. Speer, Ambiguity Resolution in Natural Language Understanding, Active Vision, Memory Retrieval, and Robot Reasoning and Actuation, in *Proc. of the IEEE International Conference on Systems, Man, and Cybernetics*, 1999, pp. 988-993.
- [5] H. Yanco and J. Drury, Classifying Human-Robot Interaction: An Updated Taxonomy, in *Proc. of the IEEE SMC 2004 International Conference on Systems, Man and Cybernetics*, 2004, pp. 2841-2846.
- [6] H. Yanco, M. Baker, B. Keyes, and P. Thoren, Analysis of Human-Robot Interaction for Urban Search and Rescue, in *Proc. of PERMIS*, 2006.
- [7] J. Casper and R. R. Murphy, Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 33, no. 3, 2003, pp. 367-385.
- [8] G. Bekey, R. Ambrose, V. Kumar, A. Sanderson, B. Wilcox, and Y. Zheng, Final Report, World Technology Evaluation Center, Inc. Panel on International Assessment of Research and Development in Robotics, 2006.
- [9] S. Tellex and D. Roy, Spatial Routines for a Simulated Speech-Controlled Vehicle, in *Proc. of the 2006 ACM Conference on Human-Robot Interaction*, 2006, pp.156-163.
- [10] C. Sidner, C. Lee, L. Morency, and C. Forlines, The Effect of Head-Nod Recognition in Human-Robot Conversation, in *Proc. of the 2006 ACM Conference on Human-Robot Interaction*, 2006, pp. 290-296.
- [11] O. Khatib, M. Peshkin, and Y. Matsuoka, pHRI - physical Human-Robot Interaction, in *Proc. of the World Technology Evaluation Center, Inc. Workshop: Review of U.S. Research in Robotics*, 2004.
- [12] J. Hertzberg and A. Saffiotti, Using semantic knowledge in robotics, *Robotics and Autonomous Systems*, vol. 56, no. 11, 2008, pp. 875-877.
- [13] G. Randelli, Improving Human-Robot Awareness through Semantic-driven Tangible Interaction, Ph.D. Thesis, Sapienza University of Rome, 2011.
- [14] J. Marciniak and Z. Vetulani, Ontology of Spatial Concepts in a Natural Language Interface for a Mobile Robot, *Applied Intelligence*, vol. 17, no. 3, 2002, pp. 271-274.

- [15] M. Skubic, C. Bailey, and G. Chronis, A Sketch Interface for Mobile Robots, in Proc. of the IEEE International Conference on Systems, Man and Cybernetics, 2003, pp. 918-924.
- [16] G. J. M. Kruijff, H. Zender, P. Jensfelt, and H. I. Christensen, Situated dialogue and spatial organization: What, where... and why, International Journal of Advanced Robotic Systems, vol. 4, no. 2, 2007, pp. 125-138.
- [17] M. Hasanuzzaman, T. Zhang, V. Ampornaramveth, H. Gotoda, Y. Shirai, and H. Ueno, Adaptive visual gesture recognition for human-robot interaction using a knowledge-based software platform, Robotics and Autonomous Systems, vol. 55, no. 8, 2007, pp. 643-657.
- [18] J. Modayil and B. Kuipers, Bootstrap learning for object discovery, in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004, pp. 742-747.
- [19] J. R. Curran, S. Clark, and J. Bos, Linguistically Motivated Large-Scale NLP with C&C and Boxer, in Proc. of the ACL 2007 Demo and Poster Sessions (Association for Computational Linguistics), 2007, pp. 33-36.
- [20] T. Matsuzaki, Y. Miyao, and J. Tsujii, Efficient HPSG parsing with supertagging and CFG-filtering, in Proc. of IJCAI-07, 2007.
- [21] R. Kaplan, S. Riezler, T. H. King, J. T. Maxwell III, A. Vasserman, and R. Crouch, Speed and accuracy in shallow and deep stochastic parsing, in Proc. of HLT and the 4th Meeting of NAACL, 2004.
- [22] S. Clark and J. R. Curran, Parsing the WSJ using CCG and log-linear models, in Proc. of ACL-04, 2004, pp. 104-111.
- [23] J. Hockenmaier, Data and Models for Statistical Parsing with Combinatory Categorical Grammar, Ph.D. thesis, University of Edinburgh, 2003.
- [24] H. Kamp and U. Reyle, From Discourse to Logic; An Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and DRT, Kluwer, Dordrecht, 1993.
- [25] A. Tyler and V. Evans, The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition, Cambridge University Press, Cambridge, 2003.
- [26] C.-R. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, and L. Prevot (eds.), Ontology and the Lexicon: A Natural Language Processing Perspective, Cambridge, 2010.
- [27] Y. Kilicaslan and E. S. Guner, Filtering Machine Translation Results with Automatically Constructed Concept Lattices, in Proc. of 8th International Conference on Concept Lattices and Their Applications (CLA 2011), 2011, pp. 59-73.
- [28] P. Krajca, J. Outrata, and V. Vychodil, Parallel Recursive Algorithm for FCA. In: Proc. CLA 2008, 2008, pp. 71-82.
- [29] H. Takeda and N. Toyoaki, Some Theoretical Considerations on Integration of Ontologies, Nara Institute of Science and Technology, Japan, 1998.
- [30] R. Jackendoff, Semantic Structures, The MIT Press, 1990.
- [31] ROS, <http://www.ros.org>