

# RoboCup@Home Spoken Corpus: Using Robotic Competitions for Gathering Datasets

Emanuele Bastianelli<sup>(†)</sup>, Luca Iocchi<sup>(‡)</sup>, Daniele Nardi<sup>(‡)</sup>,  
Giuseppe Castellucci<sup>(\*)</sup>, Danilo Croce<sup>(●)</sup>, Roberto Basili<sup>(●)</sup>

(†) DIAG

Sapienza University of Rome

{nardi,iocchi}@dis.uniroma1.it

(‡) DICII, (\*) DIE, (●) DII

University of Rome, Tor Vergata

{basili,croce}@info.uniroma2.it, {bastianelli,castellucci}@ing.uniroma2.it

**Abstract.** The definition of high quality datasets for benchmarking single components and entire systems in intelligent robots is a fundamental task for developing, testing and comparing different technical solutions. In this paper, we describe the methodology adopted for the acquisition and the creation of a spoken corpus for domestic and service robots. The corpus has been inspired by and acquired in the RoboCup@Home setting, with the involvement of RoboCup@Home participants. The annotated data set is publicly available for developing, testing and comparing speech understanding functionalities of domestic and service robots, not only for teams involved in RoboCup@Home or in other competitions, but also for research groups active in the field. We regard the construction of the dataset as a first step towards a full benchmarking methodology for spoken language interaction in service robotics.

## 1 Introduction

The creation of data sets for benchmarking different components of an intelligent robot is an important task. Suitable and high-quality data sets allow for both developing and testing new solutions and to compare existing ones. However, creating high-quality data sets is not trivial, since: 1) a proper design of data collection must be performed, depending on the tasks to be measured with the data set; 2) a proper data acquisition campaign must be executed to ensure that data will meet the requirements defined in the design phase; 3) the generation of the ground truth needed to evaluate performance of the tested modules, typically time consuming and requiring a substantial human effort. Moreover, when the data set is related to human-robot interaction issues, the additional challenge is to collect a wide and diverse data set suitably representative of different users.

In this paper, we describe the design, the collection and the generation of the corresponding ground truth for a spoken corpus to be used for developing

and testing speech recognition capabilities of domestic and service robots. In the definition of the scenario, we took inspiration from the RoboCup@Home environments and tasks. The corpus will be thus very relevant for RoboCup@Home teams. The resource is publicly available for researchers in the field at the url <http://sag.art.uniroma2.it/HuRIC.html>, in particular for RoboCup@Home teams.

The main motivation of using RoboCup@Home (and in general robotic competitions) for acquiring data sets stems from the fact that competitions provide an ideal context for benchmarking functionalities. However, this kind of benchmarking is not actually performed during a competition, because the main focus is to evaluate and compare performance of entire systems. Conversely, the ability to benchmark individual system components is needed and efforts in this direction are ongoing [1]. RoboCup@Home provided the proper context for developing the benchmark, since at the competition venue many researchers that are addressing the problems to be benchmarked can provide feedback and suggestions and guarantee the quality and the significance of the acquired data.

Summarizing, by exploiting robotic competitions, it is possible to significantly improve the quality and the significance of data sets used for benchmarking important functionalities for intelligent robots. In this paper we present an instance of this method, applied to benchmarking speech understanding capabilities of a domestic and service robot through the RoboCup@Home competition.

The paper is organized as follows. Next section describes some related work in the development of linguistic resources for speech understanding. In Section 3 we describe the design and the implementation of the acquisition process. In Section 4 we provide details about the developed corpus, by defining the type of annotations adopted on the gathered data. Finally, a discussion about the use of competitions for collecting data sets for benchmarking is provided in the final section.

## 2 Related Works

Annotated resources have always been used in the Natural Language Processing field with the aim of learning language rules from observations. Semi-automatic methods to build grammars, as well as more advanced Machine Learning based system for *POS-tagging*, *Syntactic* and *Semantic Parsing* have been realized exploiting such resources. This brought to the development of large scale annotated corpora (e.g. FrameNet [2], Penn Treebank [9], PropBank [10]) inspired by sound linguistic theories that helped in the definition of many state-of-the-art Statistical Learning approaches for NLP tasks. Even though these resources are built to be as general as possible, they do not cover all the different cases and phenomena implied by human language. As a consequence, their reuse in heterogeneous domains is not straightforward. The generalization as attempted by ML algorithms is basically biased by the employed data. Large performance drops can be noticed in out-of-domain conditions, as reported in [11, 6], where a Semantic Parsing system trained over a specific application-domain corpus shows

a significant performance drop when applied to different domains.

For these reasons, in the recent years some corpora for the automatic understanding of robot commands in Natural Language have been produced. First of all, it is important to highlight the fact that NL Human Robot Interaction deals with different aspects of language processing. Spoken interaction implies a Speech Recognition stage, while understanding the meaning of a sentence representing a command needs some form of semantic parsing. Finally, also the translation of the meaning of a sentence in the final grounded representation can be learned, and thus a resource containing all the above information is interesting.

The resources available so far have taken into account only a subset of these different aspects. For example, the work by Bugmann *et al.* [3] focuses the attention on the analysis of the semantic primitives contained in the utterances pronounced by the user in a route instruction navigation task, providing utterances paired with the related recorded audio. Kuhlmann *et al.* [8] produced a corpus of commands for the Simulator League competition @Robocup. The meaning of the sentences representing the commands is here expressed using CLang (Coach Language), a specific language that can be compiled by the simulation environment of the competition in order to change the behavior of simulated soccer players. Other resources have been gathered using crowd-sourcing to produce data with a high degree of flexibility in term of language. For example, in Tellex *et al.* [12] a corpus of written commands for navigation and manipulation tasks have been realized and exploited. Here an analysis of how the spatial domain is modelled in such commands is carried out through *Spatial Description Clauses (SDCs)*. These are semantic structures composed by a figure, a verb, a spatial relation and landmark and represent a linguistic constituent that can be grounded in the real world. Similarly, in [4] Kais presents a corpus of natural language commands for a manipulator acting in a simulated discrete 3-dimensional board. The semantic information provided is modeled through the formal Robot Command Language, encoding both semantics about actions and spatial relations between objects.

However, these corpora are highly domain or system dependent. In this context, our main aim is to build a corpus containing information that are still specific of an application domain, e.g. the house service robotics, but at the same time based or inspired by general linguistic theories. By doing this, we want to offer a level of abstraction in our resource that is independent from the robotic platform, but yet motivated by largely supported theories. Multiple *semantic theories* can be applied to describe the aspects of the world that should be taken in account by a NL HRI system. We came to the point that, for our first investigation, two main features are required: first, the robots are supposed to execute actions, possibly corresponding to a user command; second, these actions take place in a physical environment. For the first issue, we pointed out *Frame Semantics* [5] as a possible solution to model the semantics of actions. For the second, we addressed the *Holistic Spatial Semantics* [13] to model the spatial referring expressions in spoken language.

Moreover, we wanted to offer information for each step of a possible NL processing chain (e.g. Speech Recognition, NL Understanding, etc.). For these reasons, each sentence in our corpus is paired with one or more audio files. We are also working on the possibility of providing the grounded version of the command, with respect to some environment (e.g. different house settings).

### 3 The Acquisition Methodology

The dataset described in this paper has been collected in two modalities: i) by remote interaction with the Web system described in this section; ii) by interviewing members of the teams participating at the RoboCup@Home 2013 competition. In both cases, the Web portal described in this section has been used for the acquisition.

The RoboCup@Home corpus is composed by a set of utterances representing commands in a home environment. Since our aim has been to produce a complete resource for NL HRI, we provide both audio and textual representation of each gathered command. Each recorded utterance is coupled with its correct transcription, that has been checked by an operator either controlling directly the user insertion or later, during a validation phase. Users have been also requested to pronounce sentences inserted by others, so that multiple spoken versions of the same sentence are included.

In the first phase of the acquisition process, users could access the Web portal showed in Figure 1 to record the commands. General situations involved in an interaction were described in the portal by displaying text and images. Each user was asked to give a command inherent to the depicted situation. In order to provide data representing realistic conditions, a portion of the gathering took place in the competition venues and in a cafeteria, thus with different levels of background noise. Moreover, the users did not receive any constraint about what to command to the robot, except for the description of the situation. As a consequence, the uttered expressions exhibit large flexibility in lexical choices and syntactic structures, again reflecting a “realistic application” condition.

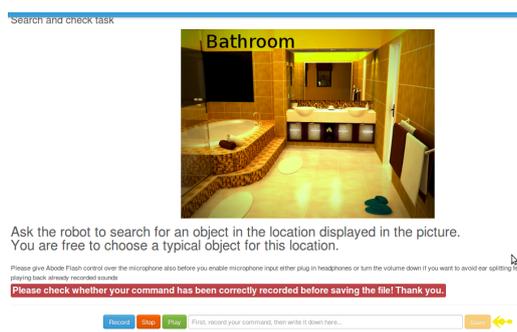


Fig. 1. The web portal used for the gathering through crowd-sourcing

In a second phase, all sentences corresponding to the transcriptions have been annotated with different syntactic and semantic information. POS-tags and syntactic dependency types have been automatically provided by the CoreNLP<sup>1</sup> system [7], and subsequently validated during the annotation process. Semantic information has been annotated according to *Frame Semantics* and *Holistic Spatial Semantics* by two expert annotators. In the last phase of the annotation process, all the tagged information has been validated by a third expert. A dedicated tool, the *Data Annotation Platform (DAP)* has been implemented and used in order to facilitate the annotation and validation process: its front-end is showed in Figure 2. This tool provides the possibility to manage linguistic information at different levels as tagging the semantics, the syntax in term of dependency types, the POS-tag and allowing to change the lemma of each word. A specific functionality of DAP allows also the user to assign a quality score to each audio file, in order to reject the one that are too noisy. In a similar way, it is possible to mark syntactically wrong sentences that have been inserted by mistake.

Other information, as speakers' generalities (e.g. age, nationality, background experience in HRI) and the specific device used for the recordings are saved together with the annotations.

The screenshot shows a web interface titled "Syntactic and Semantic Analysis". It displays a table with columns: ID, Word, Taking, Spatial\_relation, Lemma, POS, Dep\_type, and Dep\_father. Each cell in the table contains a dropdown menu with a "Remove" button. The table shows the following data:

ID	Word	Taking	Spatial_relation	Lemma	POS	Dep_type	Dep_father
1	take	Lexical_unit	----	take	VB	root	0
2	the	Theme	Trajector	the	DT	det	3
3	book	Theme	Trajector	book	NN	dobj	1
4	on	Source	Spatial_Indicator	on	IN	prep	1
5	the	Source	Landmark	the	DT	det	6
6	table	Source	Landmark	table	NN	pobj	4

Below the table is an "Execute Update" button.

Fig. 2. The Data Annotation Platform

## 4 Corpus Description

In this section an analysis of the corpus characteristics is carried out. General statistics about the composition of the corpus are here reported, as well as accurate measurements regarding the annotation process. More details are not

<sup>1</sup> [nlp.stanford.edu/software/corenlp.shtml](http://nlp.stanford.edu/software/corenlp.shtml)

provided here for lack of space, but are available on the official website of the resource (see Section 1).

#### 4.1 Corpus Statistics

As previously stated, the RoboCup@Home corpus is composed by a set of audio files representing robot commands in a home environment. Each audio file is paired with its correct transcription. These are annotated with different linguistic information: lemmas, POS-tags, dependency trees, Frame Semantics and Spatial Semantics. Table 2 reports the number of audio files together with the number of sentences corresponding to their transcriptions. In order to provide training material for ASR engines, we also asked different speakers to pronounce the same command. The average number of sentences per audio file is reported in the aforementioned Table. The recordings took place during the Robocup 2013, so speakers with different nationalities have been interviewed. Involving non native-English speakers has been a first step in the attempt of offering also training material for building non native-accent acoustic models for ASR. Table 1 reports statistics about the nationality of the different speakers.

Each user has been required to insert and record 9 commands during the acquisition process. After removing the audio files considered too noisy, an average of about 8.1 audio files per speaker has been evaluated.

Nationality	#
Australia	3
Brazil	1
UK	2
Chile	2
China	1
Cyprus	1
Czech Republic	1
Holland	5
German	4
India	1
Indonesia	1
Italy	5
Japan	1
Mexico	1
Spain	2
Syria	1
USA	4
Total	36

**Table 1.** Distribution of the nationality of the speakers

#audio files	#sentences	#audio file per sentence
292	177	~1.64

**Table 2.** Number of audio files and sentences

Imperative	Descriptive	Definitional
150	14	13

**Table 3.** Distribution of utterance classes

The situations presented to the user through the Web portal belonged to three distinct categories, each corresponding to a different pragmatics of the command. In fact, each scene required the user to pronounce either a direct command as “bring me the mug that is on the table”, or a description about the environment, e.g. “there is a bottle on the table”, or the definition of a category of a referenced entity in the scene, e.g. “this is the living room”. We then classified those sentences respectively as *imperative*, *descriptive* or *definitional*. Table 3 shows the number of sentences for each of these classes.

Statistics about the linguistic information are reported in Table 4 and 5. Table 4 reports the number of fine-grain POS-tags annotated and validated in the whole corpus, while Table 5 shows the distribution of the general coarse-grain POS-tags, e.g. verbs or nouns.

POS	#
CC	31
CD	16
DT	285
EX	11
FW	1
IN	134
JJ	43
JJS	1
MD	28
NN	365
NNS	22
POS	0
PRP	79

POS	#
PRP\$	20
RB	16
RP	1
TO	66
UH	38
VB	165
VBD	6
VBG	2
VBN	1
VBP	11
VBZ	29
WDT	3
WRB	2

POS	#
CC	31
CD	16
DT	285
EX	11
FW	1
IN	134
J	44
MD	28

POS	#
N	387
P	99
RB	16
RP	1
TO	66
UH	38
V	214
W	5

**Table 5.** Coarse-grain morpho-syntactic information

**Table 4.** Fine-grain morpho-syntactic information

## 4.2 Annotating Frame Semantics

One of the first purposes of the RoboCup@Home corpus was to provide linguistic information of different sort about natural language commands. The amount of information should enable a house service robot to completely understand their meaning.

In a house scenario, we expect mainly to have users giving commands to their “robotic butlers”. Commands are then expressions of the expectation of a user to have a robot performing the desired action. For this reason we concluded that a way of representing how actions are modeled through language was necessary to fill the gap between the linguistic knowledge about the semantics of actions and the robotic actions. We pointed out that *Frame Semantics* fitted this case. This linguistic theory generalizes the notion of action by making reference to a

situation, representing it as a *Semantic Frame* [5]. A frame is a micro-theory about a real world situation describing **actions**, such as *moving*, or more generally **events**, such as *natural phenomena* or **properties**. A set of semantic roles is associated to each frame, i.e. the descriptors of the different elements involved in the described situation (e.g. the *Goal* of a movement). Our hypothesis is that semantic frames represent a fundamental concept in NL HRI, as they can be straightforwardly linked to robot’s actions. Moreover, linguistic resources providing Frame Semantics based information have been produced over the years, as FrameNet [2].

For the RoboCup@Home corpus a subset of FrameNet-inspired semantic frames have been selected, according to the most common actions that a house robot would perform. Table 6 reports statistics about the annotated frames, together with the relative frame elements. It is worth noting that some frames have been slightly adapted with respect to their definition in FrameNet, e.g. the frame *Scrutiny* has been called *Searching*. As an example, according to the defined set of frames, in the command “*go in front of the couch*” we annotated the *Motion* frame as evoked by the verb *go*. The phrase *in front of the couch* is labeled as the GOAL frame element, representing the destination of the motion action. The instantiated frame finally encodes all the information needed to the robot to understand what action to perform, together with the arguments involved in the command, i.e. in the example above the object near which to move.

The Frame Semantic annotation process usually follows three steps. First, all the expressed actions in a sentence must be recognized: this merely means finding all the possible words evoking a frame, and associating the correct semantic frame to each of them. This process is called *Frame Prediction (FP)*. Second, given a frame, the spans (in terms of words) of the different frame elements in the sentence must be identified. We refer to this task as the *Boundary Detection (BD)* process. Finally, the correct label representing the frame element name must be associated to each span identified during the BD task, e.g. GOAL. According to the practice in the generation of annotated resources, the *Inner-Annotator-Agreement (IAA)* between the two annotators has been evaluated as a measurement of the quality of the annotations. For each of the aforementioned steps, Precision, Recall and F-Measure have been measured, considering in turn one annotator as the gold standard and evaluating the other against him. The mean of the scores of the two annotators has been finally considered as the IAA. These results are reported in Table 7. For the BD and the AC subtasks, two different measures have been reported: the *exact match* and the *token match*. The first represents the percentage of roles that have been exactly tagged, meaning that a frame element has been correctly tagged only if its entire span perfectly matches the Gold Standard one. The second measure refers to the percentage of token correctly tagged inside the labeled spans. From this Table is possible to notice how difficult is tagging the Frame Semantics, especially the BD and AC steps. Different factors biased the scores of this two steps. First a slight misalignment in the FP phase reduces it, as tagging a wrong frame compromises the further processing. Second, in some cases the annotators disagreed on the span

of some frame elements, as the CATEGORY for the *Being.in.category* frame. For example, in the command “*this is a living room with a black table*”, one annotator tended to label only the phrase *a living room* as the CATEGORY, while the other used to annotate the whole span corresponding to *a living room with a black table*.

<b><i>Attaching</i></b>	2		
ITEM	2		
<b><i>Being.in.category</i></b>	14		
CATEGORY	14		
ITEM	14		
<b><i>Being.located</i></b>	20		
LOCATION	11		
PLACE	6		
THEME	20		
<b><i>Bringing</i></b>	37		
AGENT	6		
BENEFICIARY	13		
GOAL	24		
MANNER	1		
SOURCE	9		
THEME	37		
<b><i>Change.operational.state</i></b>	3		
DEVICE	3		
OPERATIONAL_STATE	2		
<b><i>Closure</i></b>	1		
CONTAINER_PORTAL	1		
<b><i>Entering</i></b>	1		
GOAL	1		
<b><i>Following</i></b>	30		
AREA	1		
COTHEME	30		
GOAL	5		
MANNER	6		
PATH	1		
SPEED	1		
THEME	6		
		<b><i>Giving</i></b>	2
		RECIPIENT	2
		THEME	2
		<b><i>Inspecting</i></b>	3
		DESIRED_STATE	1
		GROUND	3
		INSPECTOR	1
		<b><i>Motion</i></b>	39
		AREA	1
		GOAL	38
		MANNER	1
		PATH	1
		THEME	8
		<b><i>Placing</i></b>	10
		AGENT	1
		GOAL	10
		THEME	10
		<b><i>Searching</i></b>	24
		COGNIZER	5
		GROUND	7
		PHENOMENON	24
		PURPOSE	5
		<b><i>Taking</i></b>	12
		AGENT	4
		PURPOSE	2
		SOURCE	5
		THEME	12

**Table 6.** Distribution of Frames and related Frame Elements

### 4.3 Annotating Spatial Semantics

After having found the way of linking the actions as they are represented through language and in the robot’s world, it becomes fundamental for us to consider that these agents are supposed to act in a physical environment. We then focused on how the spatial domain is modeled through language, especially in human-robot standard interactions. Even though *Frame Semantics* is able to capture

	FP			BD			AC		
	P	R	F1	P	R	F1	P	R	F1
Exact Match	95.2	95.2	95.2	84.5	84.5	84.4	82.8	82.8	82.7
Token Match	-	-	-	89.9	89.9	89.8	85.0	85.0	85.0

**Table 7.** *Frame Semantics* Inter Annotators Agreement

some of these aspects (e.g. some dynamic spatial references as the destination of a motion), we realized that in some cases the granularity level offered by this theory was not appropriate. Understanding the spatial relations holding between two or more entities can be crucial for HRI. If we consider the command “*move near the couch in the living room*”, we find out that *Frame Semantics* is not able to capture the relation holding between *the couch* and *the living room*. as the whole sequence *near the couch in the living room* is considered as the destination of the motion trajectory, i.e. the GOAL frame element. Identifying such relation would allow a robot to understand which is the couch the user is referring to, among all the couches present in the world known by the robot.

We then looked at the *Holistic Spatial Semantics* [13] to model the static spatial relation expressed in the spoken commands. This theory defines the basic concepts in the domain of natural language spatial expressions. It helps to make reference to the location or the trajectory of a motion, usually involving one referent in a discourse. It defines the concept of *spatial relation*, as a composition of different *spatial roles* present in a sentence. These can be a TRAJECTOR, i.e. the entity whose location is of relevance, a LANDMARK, i.e. the reference entity by which the location of the trajectory of the motion is fully specified, or a SPATIAL\_INDICATOR, i.e. the part of a sentence holding and characterizing the nature of the whole relation. For example, in the sentence “*go near the couch in the living room*”, the preposition “*in*” is the SPATIAL\_INDICATOR of the relation between “*table*” and “*kitchen*”, respectively a TRAJECTOR and a LANDMARK. Even though Spatial Semantics defines also other spatial roles that model dynamic spatial relations, we decide to rely only on this restricted set in order to avoid an excessive overlap with the Frame Semantics. In fact the simple meaning representation structure of a spatial relation composed by three roles perfectly suits our needing. The LANDMARK and the SPATIAL\_INDICATOR offer all the information needed to disambiguate the position of a referred entity (i.e. the TRAJECTOR), easily revealing which is respectively the reference point and the type of relation and the relation.

*Spatial Semantics* in term of these three roles have been annotated over the whole HuRIC. Table 8 reports the number of spatial relations annotated over the three datasets, together with the total number of spatial roles. It is worth noting that the number of LANDMARKS is different from the other two roles because sometimes it can be implicit, e.g. *go near [the table]<sub>TRAJECTOR</sub> [on the right]<sub>SPATIAL\_INDICATOR</sub>*. The average number of spatial relations and roles per

sentence is also reported. The *Inter-Annotator-Agreement* has been evaluated for each spatial role. It has been measured in the same way as for the *Frame Semantics*, and is reported in Table 9, considering both the *exact match* and the *token match* measures.

	#
<b><i>Spatial_relation</i></b>	47
TRAJECTOR	47
SPATIAL_INDICATOR	47
LANDMARK	41

**Table 8.** Distribution of Spatial Relations and Spatial Roles

	TRAJECTOR			SPATIAL IND.			LANDMARK		
	P	R	F1	P	R	F1	P	R	F1
Exact Match	85.8	88.6	85.7	81.4	81.4	81.3	84.7	84.7	84.6
Token Match	81.6	81.6	81.6	86.1	86.1	86.0	83.8	83.8	83.7

**Table 9.** *Spatial Semantics* Inter Annotators Agreement

## 5 Discussion

Robotic competitions have an important role for testing integrated systems and compare performance of different teams in solving complex tasks, but are also very important settings for benchmarking specific functionalities of the robot. However, these benchmarking activities are rarely performed during a competition, usually because of time constraints and of the need to test entire systems. Nonetheless, the competition setting provides for an ideal context to acquire data that can be used for subsequent benchmarking.

We thus believe that robotic competitions, and RoboCup in particular, could gain if, in parallel with running the competitions, their set-up phases could be used to acquire data sets in typically more realistic scenarios than the ones each research group can recreate in its laboratory. Indeed, acquiring data during the competitions allows for reproducing similar characteristics, such as general environmental conditions, background noise, sensors, etc.

In this paper, we have described this approach applied to the speech understanding capability of a domestic and service robot, involved in RoboCup@Home competitions. Although the competition is focused on testing entire systems, the parallel acquisition of data for subsequent benchmarking of the speech understanding module in the same scenario of the actual competition is an important task for improving performance of this capability over time.

The publicly available RoboCup@Home spoken corpus described in this paper will thus help development, test and comparison of the speech understanding capabilities of domestic and service robots, not only for RoboCup@Home teams or teams participating to some competitions, but for any research group interested in the research field.

**Acknowledgment:** authors are thankful to Cristina Giannone for her indispensable support in the development of the DAP system.

## References

1. Amigoni, F., Bonarini, A., Fontana, G., Matteucci, M., Schiaffonati, V.: Benchmarking through competitions. European Robotics Forum–Workshop on Robot Competitions: Benchmarking, Technology Transfer, and Education (2013)
2. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The berkeley framenet project. In: Proceedings of the ACL 98. pp. 86–90. Association for Computational Linguistics, Stroudsburg, PA, USA (1998)
3. Bugmann, G., Klein, E., Lauria, S., Kyriacou, T.: Corpus-based robotics: A route instruction example. In: in Proceedings of IAS-8. pp. 96–103 (2004)
4. Dukes, K.: Train robots: A dataset for natural language human-robot spatial interaction through verbal commands. In: ICSR. Embodied Communication of Goals and Intentions Workshop (October 2013)
5. Fillmore, C.J.: Frames and the semantics of understanding. *Quaderni di Semantica* 6(2), 222–254 (1985)
6. Johansson, R., Nugues, P.: The effect of syntactic representation on semantic role labeling. In: Proceedings of COLING. Manchester, UK (August 18-22 2008)
7. Klein, D., Manning, C.D.: Accurate unlexicalized parsing. In: Proceedings of ACL’03. pp. 423–430 (2003)
8. Kuhlmann, G., Stone, P., Mooney, R., Shavlik, J.: Guiding a reinforcement learner with natural language advice: Initial results in RoboCup soccer. In: The AAAI-2004 Workshop on Supervisory Control of Learning and Adaptive Systems (July 2004)
9. Marcus, M.P., Santorini, B., Marcinkiewicz, M.A.: Building a large annotated corpus of english: The penn treebank. *COMPUTATIONAL LINGUISTICS* 19(2), 313–330 (1993)
10. Palmer, M., Gildea, D., Xue, N.: Semantic Role Labeling. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers (2010)
11. Pradhan, S.S., Ward, W., Martin, J.H.: Towards robust semantic role labeling. *Comput. Linguist.* 34(2), 289–310 (2008)
12. Tellex, S., Kollar, T., Dickerson, S., Walter, M.R., Banerjee, A.G., Teller, S.J., Roy, N.: Understanding natural language commands for robotic navigation and mobile manipulation. In: Burgard, W., Roth, D. (eds.) AAAI. AAAI Press (2011)
13. Zlatev, J.: Spatial semantics. *Handbook of Cognitive Linguistics* pp. 318–350 (2007)