Inferring Robot Actions from Verbal Commands Using Shallow Semantic Parsing

A. Sutherland, S. Bensch, and T. Hellström
Department of Computing Science, Umeå University, Umeå, Sweden

Abstract— *Efficient and effective speech understanding systems are highly interesting for development of robots working together with humans. In this paper we focus on interpretation of commands given to a robot by a human. The robot is assumed to be equipped with a number of pre-defined action primitives, and an uttered command is mapped to one of these actions and to suitable values for its associated parameters. The approach taken is to use data from shallow semantic parsing to infer both the action and the parameters. We use labeled training data comprising sentences paired with expected robot action. Our approach builds on the hypothesis that the expected action can be inferred from semantic frames and semantic roles, information that we retrieve from the Semafor system. The generated frame names and semantic roles are used to learn mappings to expected robot actions and their associated parameters. The results show an accuracy of 88% for inference of action alone, and 68% for combined inference of an action and its associated parameters. Given the large linguistic variety of the input sentences, and the limited size of data used in for learning, these results are surprisingly positive and promising.*

Keywords: HRI, NLP, Robotics, Semantics, Arbitration

1. Introduction

Speech is one of the most efficient means of communication for humans, and has also been extensively addressed in human-robot interaction research [1], [2]. While robust speech recognition is a major unsolved problem in natural language processing (NLP), challenges also remain in other areas of NLP, such as syntactic and semantic analysis. Even if these problems would be solved, a general method to generate correct robot responses to speech requires a level of intelligence that is out of reach for current research in both cognitive science and artificial intelligence. The challenges in finding a general solution has of course not prevented researchers from proposing solutions to specific domains and sentence structures. Most implementations of NLP in robotics is concerned with imperative commands and this is also the target for the work presented in this paper. We propose a method to create mappings from sentences to expected robot actions. Humans, not least young children, are able to the perform such learning in a non supervised manner, i.e. without being explicitly told what to do when a certain sentence is uttered. For an excellent analysis of this process see [3]. While still awaiting human level intelligence in robots, we make the task somewhat simpler by providing the expected robot action for each sentence. Thus, we provide labeled training data comprising sentences paired with the expected robot action. A robot action comprises the name of the pre-defined action, and values for one or several parameters specifying the action. Our approach build on a hypothesis that the expected action can be inferred from shallow semantic data. In the learning phase, the labeled sentences are semantically parsed using the commonly available Semafor system [4]. The generated frame names and semantic roles are used to create mappings to expected robot actions including associated parameters. The evaluation shows very good results for cross-validated data.

This paper is organized as follows: Section 2 gives a brief overview of earlier related work. The theory behind semantic roles is briefly described in Section 3, followed by a description of our approach for generation of mappings from sentences to frames and semantic roles. The mechanism for inference of actions and parameters is described in Section 4. Results are presented in Section 5, followed by a discussion of results and limitations in Section 6, and plans for future work in Section 7.

2. Related earlier work

Substantial research has been focused on speech-based systems for giving commands to robots. In [5], a system for programming a robot for complex tasks through verbal commands is presented. The system filters out unknown words, maps predefined keywords to actions and parameters, and generates graphs representing the required task. The authors in [6] propose a speech-based robot system for controlling a robotic forklift. The proposed system requires commands to be given according to a given syntax. In [7], teaching soccer skills via spoken language is addressed. The vocabulary is predefined and focus is rather on constructing advanced control expressions than on language understanding. The authors in [8] use labeled sentences to learn a combinatory categorical grammar (CCG) for the specific task of interpretation of route instructions. In [9], specific techniques for incremental language processing coupled to inference of expected robot actions are described. The approach is

to construct a grammar that facilitates incremental analysis such that robots can act pro-actively already during a verbal command is given.

The present work is similar to the once mentioned above in the aim to interpret natural language sentences by mapping sentences to expected robot actions. However, it differs by the method of using semantic frames and roles from an existing parser as inputs in the mapping.

Other attempts to human-robot interaction through natural language build on more traditional grammatical analysis combined with reasoning mechanisms to generate suitable robot actions. Low recognition rates and ungrammatical, incomplete or fragmentary utterances have been addressed in several ways. The authors in [10] constrain the task and use incremental language analysis based on CCG, regular expression-based filter and a trigram statistical model to process fragmentary or otherwise ungrammatical utterances.

3. Shallow semantic parsing

Shallow semantic parsing, also called semantic-role-labeling, is the task of finding semantic roles for a given sentence. Semantic roles describe general relations between predicates and its arguments in a sentence. For example, in a sentence like "Mary gave the ball to Peter", "gave" is the predicate, "Mary" represents the semantic role *donor*, "the ball" represents the semantic role *theme*, and "Peter" represents the semantic role *recipient*.

FrameNet [11] is a system with large amounts of such analyzes for English sentences. Whereas other attempts, like *PropBank* [12], assign roles to individual verbs, FrameNet assign roles to *frames*. A semantic frame includes a list of associated words and phrases that can potentially evoke the frame. Each frame also defines several semantic roles corresponding to aspects of the scenario represented by the frame. The Semafor system (Semantic Analyzer of Frame Representations) is built on FrameNet and provides both an on-line service (http://demo.ark.cs.cmu.edu/parse) and downloadable code for off-line use of the system. Semafor is reported [13] to have achieved the best published results up to 2012 on the *SemEval 2007 frame-semantic structure extraction task* [14]. In the present work we use the Semafor on-line system for extraction of frames and semantic roles from all sentences used in the experiments.

4. Description of method

We propose a method by which the expected actions for a verbally uttered commands to a robot can be learned, such that the robot automatically can determine what to do when hearing a new sentence. The robot learns how to infer action and parameters from a set of labeled example sentences. Each sentence is parsed by a shallow semantic parser that produces frames and associated semantic roles. If multiple frames occur, the frame related to the predicate is selected, and denoted as the *primary* frame. Conditional probabilities for how these entities relate to expected actions and associated parameters are estimated and used to construct the necessary inference mechanisms.

The example sentences used in this paper were manually generated. Each sentence was labeled with one of n_A robot actions $a_1, ..., a_{n_A}$ and m_a associated parameters $p_1, ..., p_{m_a}$ (see Table 1). A total of 94 sentences representing plausible commands that a human might utter to a robot were generated. Some examples are given in Table 4.

Table 1: Pre-programmed robot actions a_i with associated parameters p_1, p_2.

i	a_i	p_1	p_2	Expected function
1	BRING	object	recipient	Fetches object
2	TELL	message	recipient	Relays a message
3	COLLECT	object	source	Gathers objects
4	MOVE	location		Moves self to location
5	PUT	object	location	Places an object

For the purpose of this paper, the actions did not have to be physically implemented but would in a complete system be pre-programmed in the robot.

The proposed method comprises a learning part and an inference part, as described in the following two subsections.

4.1 Learning

In the learning phase, each sentence in a training data set comprising N sentences was presented to the Semafor system, which output frames and associated semantic roles. If several frames were generated, the primary frame is selected. For our entire data set, $n_F = 21$ distinct primary frames $f_1, ..., f_{n_F}$, were generated, and are listed in Table 2 together with some of their most common semantic roles.

The proposed method builds on the hypothesis that the expected action for a command can be inferred from the primary frame of the command. To initially test this hypothesis, statistics for combinations of primary frames and labeled actions for all sentences were generated, and is presented in Table 3. The number of occurrences for each frame/action combination is shown, followed by the relative frequency after the / symbol. Most rows contain only one non-zero entry, thus supporting the hypothesis that the expected action can be inferred from the frame. However, some frames occur for more than one action, and many actions occur for several frames.

In order to infer expected action from the primary frame of a sentence, the conditional probability

$$P(Action = a_i | Frame = f_j), \quad (1)$$

i.e. for the expected action to be a_i, given a primary frame f_j, are estimated. With simplified notation and by using the

Table 2: Frame names generated by the Semafor system for the sentences used in the experiments.

i	Frame f_i	Common semantic roles
1	BRINGING	Theme Goal Source Path
2	GETTING	Event Experiencer Focal participant
3	GIVING	Donor Recipient Theme
4	NEEDING	Cognizer Dependant Requirement
5	DESIRING	Event Experiencer
6	TELLING	Addressee Message Speaker
7	STATEMENT	Message Speaker Medium
8	POLITICAL LOCALES	Locale
9	BEING NAMED	Entity Name
10	TEXT	Text Author
11	COME TOGETHER	Individuals
12	AMASSING	Mass Theme Recipient
13	GATHERING UP	Agent Individuals
14	PLACING	Agent Goal Theme
15	MOTION	Path Goal Theme
16	GRANT PERMISSION	Grantee Grantor Action
17	DEPARTING	Source Theme
18	STIMULUS FOCUS	Stimulus
19	HAVE AS REQ.	Dependant Required entity Requirement
20	LOCALE BY USE	Locale Use
21	COMPLIANCE	Act Norm Protagonist

definition of conditional probability, (1) can be written as

$$P(a_i|f_j) = P(a_i, f_j)/P(f_j), \tag{2}$$

which can be estimated from data by

$$\widehat{P}(a_i, f_j) = \#(a_i, f_j)/N \tag{3}$$

and

$$\widehat{P}(f_j) = \#(f_j)/N, \tag{4}$$

where $\#(a_i, f_j)$ denotes the total number of sentences in the training data that were labeled with action a_i and for which Semafor determines f_j as primary frame[1]. Hence, $P(a_i|f_j)$ can be estimated by

$$\widehat{P}(a_i|f_j) = \#(a_i, f_j)/\#(f_j). \tag{5}$$

The n_F different frames that appear in our scenario have in total n_R distinct associated semantic roles with the following names: *Goal, Theme, Source, Recipient, Requirement, Cognizer, Event, Experiencer, Addressee, Message, Name, Text, Donor, Individuals, Mass theme, Path, Grantee, Action, Direction,* and *Dependent*. These semantic roles are in the following denoted $r_1, ..., r_{n_R}$.

Normally, each frame only has a few semantic roles defined. When parsing an input sentence s, Semafor assigns substrings of s as values to these semantic roles.

According to the suggested approach, parameters for each robot action are related to specific semantic roles. Since the manual identification of parameters in the labeling of sentences not necessarily works by the same principles as the identification of semantic roles in Semafor, a parameter p_i is regarded as *matching* (denoted by the symbol $\sim$) a semantic role r_j if p_i is a nonempty substring of the value of r_j:

$$p_i \sim r_j \equiv p_i \text{ is a nonempty substring of the value of } r_j. \tag{6}$$

Example: Assume that the sentence "Give me the glass" is labeled with action a_1 (i.e. BRING) and parameter $p_1 =$ "glass". Semafor generates a primary frame f_3 (i.e. GIVING), and semantic role r_2 (i.e. Theme) is assigned the value "the glass" for the sentence. Hence, $p_1 \sim r_2$.

In the next section we will construct a classifier to infer expected action a_E for a sentence with a primary frame name f_E. To infer parameters for a_E, we need to estimate the probability that a parameter p_i for a_E matches a semantic role r_j, given that the primary frame is f_E (separate estimates for each $p_i, i = 1, ..., m_a$). With the introduced notation, and by using the definition of conditional probability, this can be written as:

$$P(p_i \sim r_j|f_E) = P(p_i \sim r_j, f_E)/P(f_E). \tag{7}$$

The probabilities on the right-hand-side of (7) can be estimated as follows.

$$\widehat{P}(p_i \sim r_j, f_E) = \#(f_E, p_i \sim r_j)/N \tag{8}$$

and

$$\widehat{P}(f) = \#(f_E)/N \tag{9}$$

where $\#(f_E, p_i \sim r_j)$ denotes the total number of sentences in the training data for which Semafor determines a primary frame f_E and a semantic role r_j, and the sentence was labeled with parameter p_i, satisfying $p_i \sim r_j$. The entity $\#(f_E)$ is the total number of sentences in the training data for which Semafor determines a primary frame f_E. Combining (7−9), yields the following estimation:

$$\widehat{P}(p_i \sim r_j|f_E) = \#(f_E, p_i \sim r_j)/\#(f_E). \tag{10}$$

As described in the next section, the estimated conditional probabilities are used to infer expected action and associated parameters for a given sentence.

4.2 Inference of expected action and parameters

A Bayes classifier is used to infer the expected action a_E for a sentence with a primary frame name f_E and semantic roles $r_i, i = 1, ..., n_R$. It works by inferring the action with highest conditional probability, as given by (1−5):

$$\begin{aligned} a_E &= \arg \max_{1 \le i \le n_A} \widehat{P}(Action = a_i|Frame = f_E) \\ &= \arg \max_{1 \le i \le n_A} \#(a_i, f_E)/\#(f_E) \\ &= \arg \max_{1 \le i \le n_A} \#(a_i, f_E). \end{aligned} \tag{11}$$

[1]In general, the function # denotes the number of observations for with the conjunction of all arguments are true. We simplify the notation as when we denote probabilities, and write for instance a_i instead of *Action*$= a_i$.

Frame \ Labeled Action	BRING	TELL	COLLECT	MOVE	PUT
1 Bringing	7/100%	0/0%	0/0%	0/0%	0/0%
2 Getting	4/100%	0/0%	0/0%	0/0%	0/0%
3 Giving	3/60%	2/40%	0/0%	0/0%	0/0%
4 Needing	2/100%	0/0%	0/0%	0/0%	0/0%
5 Desiring	4/100%	0/0%	0/0%	0/0%	0/0%
6 Telling	0/0%	8/100%	0/0%	0/0%	0/0%
7 Statement	0/0%	8/100%	0/0%	0/0%	0/0%
8 Political locales	0/0%	0/0%	0/0%	0/0%	0/0%
9 Being named	0/0%	0/0%	0/0%	0/0%	0/0%
10 Text	0/0%	1/100%	0/0%	0/0%	0/0%
11 Come together	0/0%	0/0%	4/100%	0/0%	0/0%
12 Amassing	0/0%	0/0%	4/100%	0/0%	0/0%
13 Gathering up	0/0%	0/0%	6/100%	0/0%	0/0%
14 Placing	0/0%	0/0%	2/11%	0/0%	17/89%
15 Motion	0/0%	0/0%	0/0%	14/82%	3/18%
16 Grant permission	0/0%	0/0%	0/0%	0/0%	0/0%
17 Departing	0/0%	0/0%	0/0%	1/100%	0/0%
18 Stimulus focus	0/0%	0/0%	0/0%	0/0%	0/0%
19 Have as requirement	0/0%	0/0%	0/0%	0/0%	2/100%
20 Locale by use	0/0%	0/0%	0/0%	0/0%	1/100%
21 Compliance	0/0%	0/0%	0/0%	0/0%	1/100%

Table 3: Occurrences/frequencies for combinations of primary frames and labeled actions, for the input data used in the experiments. Most rows contains only one non-zero entry, thus supporting the hypothesis that the expected action can be inferred from the frame.

Each one of the parameters $p_i^E, i = 1, ..., m_{a_E}$ required by action a_E is assigned the value of one of the semantic roles $r_i, i = 1, ..., n_R$ for the sentence. The procedure for inference of parameters follows the same principles as for inference of action in (7−10), and parameter values are assigned as follows:

$$p_i^E = r_{opt}, \tag{12}$$

where

$$\begin{aligned} opt &= \arg \max_{1 \le j \le n_R} \widehat{P}(p_i \sim r_j | f_E) \\ &= \arg \max_{1 \le j \le n_R} \#(f_E, p_i \sim r_j)/\#(f_E) \\ &= \arg \max_{1 \le j \le n_R} \#(f_E, p_i \sim r_j). \end{aligned} \tag{13}$$

The inference of expected action and parameters, as described above, is expressed as pseudo-code in Algorithm 1. In steps 5-6, Semafor is used to compute primary frame and semantic role values for the input sentence s. The subset of training sentences with the same primary frame is selected in step 7, such that the computation of the expected action in step 8 corresponds to (11). Values for the parameters p_i^E are computed in steps 11-12, corresponding to (12−13). The algorithm was implemented and evaluated with cross-validation, as described in the next section.

Algorithm 1 Infer expected action a_E and associated parameters p_i^E for an input sentence s.

1: **return** a_E and $p_i^E, ..., p_{m_{a_E}}^E$
2: **inputs:**
3: s : **sentence to be analyzed**
4: A : **set of training sentences labeled with action** a **and parameters** $p_1, ...p_{m_a}$
5: $f_E \leftarrow$ **the primary frame of** s
6: $r_1^E, ..., r_{n_R}^E \leftarrow$ **semantic roles for** s
7: $B \leftarrow$ **the subset of** A **with** f_E **as primary frame**
8: $a_E \leftarrow$ **the most common action** a **in** B
9:
10: **for** $i = 1$ **to** m_{a_E} **do**
11: **find the index** opt **for which** $p_i \sim r_{opt}$ **in most sentences in** B
12: $p_i^E \leftarrow r_{opt}^E$
13: **end for**

4.3 Evaluation

The developed system was evaluated using the full data set of 94 sentences. Evaluation was done by leave-one-out cross-validation, i.e. one sentence was left out of the training data set, and a model was constructed as described in Section 4.1. The model was evaluated by inferring expected

Table 4: Examples of sentences used for training and evaluation. Each sentence is labeled with expected action and associated parameter(s).

	Sentence	Expected action	p_1	p_2
1	move the chairs to the kitchen	PUT	chairs	the kitchen
2	Move 2 meters to the left	MOVE	2 meters	to the left
3	I want a glass of water.	BRING	a glass of water	
4	Robot, tell Ola the name of the book.	TELL	the name of the book	Ola
5	stash the balls in the wardrobe.	PUT	the balls	in the wardrobe
6	package all glasses into nice parcels.	PUT	all glasses	into nice parcels
7	Gather all the green balls.	COLLECT	all the green balls	
8	Robot, tell Ola the color of the ball.	TELL	the color of the ball	Ola
9	Gather dust in the room.	COLLECT	dust	in the room
10	Go to the tire storage.	MOVE	the tire storage	
11	Robot, tell the direction of the exit to me.	TELL	the direction of the exit	me
12	Bring Ola's book to me.	BRING	Ola's book	me

Table 5: Semantic parses of the sentences in Table 4, as given by the Semafor system. The table shows primary frame name and some of the generated semantic roles for the frame.

	Primary frame	Semantic role/value	Semantic role/value	Semantic role/value
1	MOTION	Theme/the chairs	Goal/to the kitchen	
2	MOTION	Theme/2 meters	Goal/to the left	
3	DESIRING	Experiencer/I	Event/a glass of water	
4	TELLING	Speaker/Robot	Addresse/Ola	Message/the name of the book
5	PLACING	Theme/the balls	Goal/in the wardrobe	
6	PLACING	Theme/all glasses	Goal/into nice parcels	
7	COME TOGETHER	Individuals/all the green balls		
8	TELLING	Speaker/Robot	Addresses/Ola	Message/the color of the ball
9	COME TOGETHER			
10	MOTION	Goal/to the tire storage		
11	TELLING	Speaker/Robot	Addresses/to me	Message/the direction of the exit
12	BRINGING	Theme/Ola's book	Goal/to me	

action and parameters for the held out sentence, as described in Section 4.2, and the procedure was then repeated 93 times such that all sentences were left out once from the training. Performance figures were computed as the average performance for all 94 training/evaluation sessions.

5. Results

In order for a robot to be able act correctly on an uttered sentence, both action and parameters have to be correctly inferred. We present results for both these tasks in Table 6. Each row in the confusion matrix shows how sentences with a specific labeled action leads to inference of various actions, shown in separate columns. Cases where no inference of action was possible are shown in the column labeled "?". At the end of each row, the accuracy for combined action and parameter inference is shown. E.g., sentences labeled with the TELL action leads in 2 cases (11%) to an incorrectly inferred BRING action, and in 16 cases (84%) to a correctly inferred TELL action. For one sentence labeled with a TELL action, no action could be inferred. The reason was that the primary frame (the TEXT frame) for this sentence occurred only once in the whole data set (see Table 3), and hence not at all in the training set for that specific sentence. Hence, no inference was possible for that sentence. The combined inference of both action and parameters, for all sentences labeled with a TELL action, was correct in 14 cases (74%). As a whole, the non-zero entries are gathered on the diagonal, which means that the inferred actions equals the labeled actions. The average accuracy for all sentences for inference of action was 88%, and for combined inference of action and parameters 68%.

6. Discussion

The proposed method builds on the hypothesis that expected actions can be inferred from shallow semantic information. We conclude that the hypothesis was valid for more than 88% of the tested sentences. Expected actions and parameters were correctly inferred for 68% of the cases.

Table 6: Confusion matrix showing number of cases/percentages for inference of expected robot actions. Figures for inference of both actions and associated parameters is shown in the right-most column. Each row contains results for sentences with one specific labeled action.

Labeled \ Inferred	BRING	TELL	COLLECT	MOVE	PUT	?	Accuracy
BRING	20/100%	0/0%	0/0%	0/0%	0/0%	0/0%	14/70%
TELL	2/11%	16/84%	0/0%	0/0%	0/0%	1/5%	14/74%
COLLECT	0/0%	0/0%	14/88%	0/0%	2/13%	0/0%	10/62%
MOVE	0/0%	0/0%	0/0%	14/93%	0/0%	1/7%	9/60%
PUT	0/0%	0/0%	0/0%	3/13%	19/79%	2/8%	17/71%

Given the large variety of sentences, and the small data set being used, the result is considered both surprising and promising. Better results can be expected by adding more data. One specific problem with limited data was discussed in the previous section: if a frame occurs only once in the data set, it is not possible to infer the expected action for that sentence since it is removed as part of the cross-validation process. Extending the data such that there are at least two sentences for each frame name, would clearly improve performance. By removing the four sentences for which the situation occurs in our data, the accuracy for inference of action improves to 92%, and for combined inference of actions and parameters to 71%.

The proposed method relies, to a very large extent, on the quality of the semantic labeling, which in our case was performed by the Semafor system. While identified frames and semantic roles do not necessarily have to be linguistically "correct", they should be consistent in the sense that semantically similar sentences should give the same results. This is unfortunately not the case with the online version of Semafor that we have been using (the downloadable version behaves somewhat differently but not better in this respect). Not only does it fail in the sense described above, but also by producing vastly different results depending on capitalization of the first letter in the sentence, and on whether the sentence is ended by a period or not. As an example, adding a period to sentence 1 in Table 4 results in a replacement of the semantic role *Goal* with *Building subparts* (also see Table 5). Another example is the sentence "Bring Mary the cup.", with varying results depending on both punctuation and replacement of "Mary" by "me". Due to such experienced problems with the Semafor system, the sentences used in the reported experiments were manually selected to ensure that the automatic semantic analysis was reasonable and consistent. This is clearly a concern for practical usage and continued research on the proposed method, but was outside the scope of the present work.

7. Future work

Since the results of the proposed method depends heavily on the quality of the semantic parsing, alternative approaches in which syntax and semantics are treated simultaneously [15] will be investigated.

As part of the inference process, the parameters for an expected action are bound to the values of certain semantic roles (12). These values are substrings like "the green ball", and "all my books" and have in this work not been further analyzed, but rather assumed to be properly interpreted by the pre-programmed action routines. This is definitely not a trivial task and contains several hard problems. The parameters are typically noun phrases, that have to be semantically analyzed and grounded to objects that the robot can perceive. This task will be a major and important part of the continued work.

References

[1] G. Bugmann, J. Wolf, and P. Robinson, "The impact of spoken interfaces on the design of service robots," *Industrial Robot: An International Journal*, vol. 32, no. 6, pp. 499–504, 2005.

[2] M. Scheutz, P. Schermerhorn, J. Kramer, and D. Anderson, "First steps toward natural human-like hri," *Autonomous Robots*, vol. 22, no. 4, pp. 411–423, 2007.

[3] J. M. Siskind, "A computational study of cross-situational techniques for learning word-to-meaning mappings," *Cognition*, vol. 61, no. 1, pp. 39–91, 1996.

[4] D. Das, N. Schneider, D. Chen, and N. A. Smith, "Probabilistic frame-semantic parsing," in *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. Association for Computational Linguistics, 2010, pp. 948–956.

[5] C. Meriçli, S. D. Klee, J. Paparian, and M. Veloso, "An interactive approach for situated task teaching through verbal instructions," in *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[6] E. Chuangsuwanich, S. Cyphers, J. Glass, and S. Teller, "Spoken command of large mobile robots in outdoor environments," in *Spoken Language Technology Workshop (SLT), 2010 IEEE*. IEEE, 2010, pp. 306–311.

[7] A. Weitzenfeld, A. Ejnioui, and P. Dominey, "Human robot interaction: Coaching to play soccer via spoken-language," *IEEE/RAS Humanoids*, vol. 10, 2010.

[8] C. Matuszek, E. Herbst, L. Zettlemoyer, and D. Fox, "Learning to parse natural language commands to a robot control system," in *Experimental Robotics*. Springer, 2013, pp. 403–415.

[9] S. Bensch and T. Hellström, "Towards proactive robot behavior based on incremental language analysis," in *Proceedings of the 2014 Workshop on Multimodal, Multi-Party, Real-World Human-Robot Interaction*. ACM, 2014, pp. 21–22.

[10] R. Cantrell, M. Scheutz, P. Schermerhorn, and X. Wu, "Robust spoken instruction understanding for HRI," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*. IEEE, 2010, pp. 275–282.

[11] C. J. Fillmore, C. R. Johnson, and M. R. Petruck, "Background to FrameNet," *International Journal of Lexicography*, vol. 16, no. 3, pp. 235–250, 2003.
[12] P. Kingsbury and M. Palmer, "From TreeBank to PropBank," in *LREC*. Citeseer, 2002.
[13] D. Das, D. Chen, A. F. Martins, N. Schneider, and N. A. Smith, "Frame-semantic parsing," *Computational Linguistics*, vol. 40, no. 1, pp. 9–56, 2014.
[14] C. Baker, M. Ellsworth, and K. Erk, "SemEval'07 task 19: frame semantic structure extraction," in *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 2007, pp. 99–104.
[15] S. Bensch, F. Drewes, H. Jürgensen, and B. van der Merwe, "Graph transformation for incremental natural language analysis," *Theoretical Computer Science*, vol. 531, pp. 1–25, 2014.