Dialogue with Robots to Support Symbiotic Autonomy

Andrea Vanzo, Danilo Croce, Emanuele Bastianelli, Guglielmo Gemignani, Roberto Basili and Daniele Nardi

Abstract Service Robotics is finding solutions to enable effective interaction with users. Among the several issues, the need of adapting robots to the way humans usually communicate is becoming a key and challenging task. In this context the design of robots that understand and reply in Natural Language plays a central role, especially when interactions involve untrained users. In particular, this is even more stressed in the framework of Symbiotic Autonomy, where an interaction is always required for the robot to accomplish a given task. In this article, we propose a framework to model dialogues with robotic platforms, enabling effective and natural dialogic interactions. The framework relies on well-known theories as well as on perceptually informed spoken language understanding processors, giving rise to interactions that are tightly bound to the operating scenario.

Keywords Dialogue modeling · Symbiotic autonomy · Mobile service robots

A. Vanzo (✉) · G. Gemignani · D. Nardi
Department of Computer, Control and Management Engineering "Antonio Ruberti",
Sapienza University of Rome, Rome, Italy
e-mail: vanzo@dis.uniroma1.it

G. Gemignani
e-mail: gemignani@dis.uniroma1.it

D. Nardi
e-mail: nardi@dis.uniroma1.it

D. Croce · R. Basili
Department of Enterprise Engineering, University of Roma Tor Vergata, Rome, Italy
e-mail: croce@info.uniroma2.it

R. Basili
e-mail: basili@info.uniroma2.it

E. Bastianelli
Department of Civil Engineering and Computer Science Engineering,
University of Roma Tor Vergata, Rome, Italy
e-mail: bastianelli@ing.uniroma2.it

1   Introduction

Robots are expected to support human activities in everyday environments, interacting with different kinds of users. In particular, domestic robots (i.e. robots operating in our homes) have already entered the market, e.g. cleaning robots or telepresence robots for elderly care. In these contexts, the interaction with the user plays a key role.

The current development of robotics technology is facing several difficulties in providing general solutions to this problem. The major causes that withhold the realization of a robust Natural Language interface consist of the enormous variety of environments, involved users and tasks to be executed, aspects that need to be understood by a robot. On the one hand, the perception capabilities of the robots make it difficult to build rich and reliable representations of the operational environment; on the other hand, combining motion and manipulation capabilities on a single platform is still very expensive and makes the size of the robot not well suited for operation in homes. While these difficulties may require some time before satisfactory solutions became available, a number of researchers are proposing to exploit Human-Robot Interaction (HRI) to enable the robot to understand the environment and accomplish tasks that would be otherwise unachievable. This line of research has been termed *Symbiotic Autonomy* [1] and it substantially relies on spoken dialogue between robots and users.

In fact, given the recent advancements in *Spoken Language Recognition and Understanding*, dialogue in Natural Language will be a major component of robotic interfaces, also considering that it will certainly be coupled with other multi-modal communication channels. In this respect, "Dialogue with Robots" has been the focus of recent research, as confirmed by special issues dedicated by several journals to this topic [2].

In the context of HRI, the focus is usually on situated dialogues. In situated dialogues, robots and humans have different representations of the shared environment, because of their mismatched perceptual capabilities. Accordingly, understanding dialogue is about more than just understanding the speech signal, words, or the utterance. Hence, for a robot that is expected to understand dialogue when talking with a human, it is essential to interpret how that dialogue relates and refers to the surrounding world.

The aim of the present work is twofold. First, we identify different HRI scenarios and situations where dialogue can be beneficial and plays a key role. As an example, dialogic interactions allow to fulfill missing information when a command has not been completely understood by the robot as well as when the resulting interpretation involves manifold ambiguities. Second, we provide pragmatic solutions to deal with this problem, along with possible frameworks that enable an effective interaction between humans and robots.

The rest of the article is structured as follows. Section 2 reports related works, while Sect. 3 describes backgrounds and proposed solutions. In Sect. 4, we identify some use cases in the context of HRI. Finally, Sect. 5 provides final remarks.

2   Related Work

In the context of Human-Robot Interaction (HRI), Natural Language Understanding has been studied starting from [3], where the focus was on a system able to process NL instructions to perform actions in a virtual environment. In the Robotic field, speech-based approaches have been applied to deploy robotic platforms in a wide plethora of environments. These techniques have been used in manipulation tasks [4], and for wheeled platforms [5, 6]. Dialogue has also been employed to instruct robots to accomplish a given unknown task, such as giving a tour [7], delivering objects [8], or manipulating them [9]. Other related works have combined speech-based approaches with other types of interactions [6, 10].

More recently, several domain-specific systems that allow users to instruct robots through Natural Language have been presented in literature. For example, in [11, 12], the authors present different methods for following Natural Language route instructions, by decoupling the semantic parsing problem from the grounding problem. In these works, the input sentences are first translated to intermediate representations, which are then grounded into the available knowledge. In [13], the authors present a preliminary version of a cascade of reusable Natural Language Processing (NLP) modules, that can be adapted to changing operational scenarios, through trainable statistical models for which HRI specific learning algorithms. These modules range from ASR re-ranking functions (e.g. [14, 15]) up to techniques to ground entities according to lexical references [14, 16]. A further refinement of such a cascade has been proposed in [17], where a standard pipeline for semantic parsing is extended toward a form of perceptually informed NLP, by combining discriminative learning, distributional semantics and perceptual knowledge. In [18], the authors show how to enable Natural Language interactions in a scenario of collaborative human-robot tasks, by mining past interactions between humans in online multiplayer games.

In [19], the authors present a probabilistic approach able to learn referring expressions for robot primitives and physical locations in a map, by exploiting the dialogue with the user. The problem of Referring Expressions Generation (REG) has also been taken into account by in [20]. They propose a hypergraph-based approach to account for group-based spatial relations and uncertainties in perceiving the environment, in the context of situated dialogues. A further refinement of their approach is introduced in [21]. Here, they develop two collaborative models for REG. Both models, instead of generating a single referring expression to describe a target object as in the previous work, generate multiple small expressions that lead to the target object with the goal of minimizing the collaborative effort. A study examining the generation of noun phrases within a spoken dialogue agent for a navigation domain is presented in [22]. Here the noun phrase generation is driven by both the dialogue history and spatial context features, e.g. view angle of the agent, distance from the target referent and the number of similar items in view. In [23], the authors

present a Natural Language generation approach which models, exploits, and manipulates the non-linguistic context in situated communication. The proposed method for the generation of referring expressions is tightly integrated with the syntactic realization of the sentence.

The problem of tackling the vocabulary in conversational systems has been addressed in [24]. They propose approaches that incorporate user language behavior, domain knowledge, and conversation context in word acquisition, evaluating such methods in the context of situated dialogue in a virtual world. In [25], the authors present four case studies of implementing a typical HRI scenario with different state-of-the-art dialogue frameworks with the goal to identify pitfalls and potential remedies for dialogue modeling on robots. They show that none of the investigated frameworks overcomes all problems in one solution. In [26], the authors focused on recovery from *situated grounding problems*, a type of miscommunication that occurs when an agent fails to uniquely map a person's instructions to its surroundings.

NLP in Robotics can be coupled with other communication channels. In [27], a flexible dialogue-based robotic system for humanlike interaction is proposed. In particular, they focus on task-based dialogues, where the robot behavior is changed based on a tight integration between Natural Language and action execution. An algorithmic framework, Continual Collaborative Planning (CCP), for modeling the integration of the different channels in situated dialogues has been proposed in [28]. This framework allows to integrate planning, acting and perception with communication. Similarly, in [29] the authors propose information-state dialogue management models for the situated domain. Here, the dialogue management model fuses information-state update theory, with a light-weight rational agency model.

Nevertheless, all the presented works are not able to recover when manifold ambiguities and missing information are found and to incrementally enhance their Natural Language Understanding from the continuous interaction with the user. Moreover, the state of the robot is often neglected, giving rise to additional ambiguities and misunderstandings. In the next section all these aspects are addressed in detail.

3   A Pragmatic Approach for Dialogue Modeling

According to the Symbiotic Autonomy paradigm, we investigated several realms, where dialogic interactions between a user and a robot are beneficial: from the robot perspective, they allow to better understand the user needs while from the user perspective, it is the most natural way to support the robot in a better comprehension of the user's requests. Section 3.1 provides some of the motivations of this work. A possible dialogue-based framework for Human-Robot Interaction is presented in Sect. 3.2.

3.1 Background and Motivation

In our earlier research on Human-Robot Interaction, we addressed the **Spoken Language Understanding** (SLU) task for the automatic interpretation of commands. Given a spoken command, this process aims at automatically analyzing the user's utterances to derive computational structures that (i) reflect the meaning of the commands and (ii) activate the robot plans. Nevertheless, the correct interpretation of a command does not merely depend only on the linguistic information that is derivable from the utterance. As suggested in [17] the SLU process does depend also on other factors, e.g., the environment surrounding the robot. As an example, the command *Take the book on the table* requires the robot to *Take* the book from the table only if there is actually a book on it; on the contrary, the same command requires the robot to *Bring* the book over the table.[1] Dialogue is crucial in order to support a proper comprehension of a command, e.g., when some information is missing. A command such as *Take the book* cannot be executed if the robot is unaware of the position of the book. In these cases, the robot could require some additional information to complete the task and fulfill the user needs.

Secondly, we analyzed the process called **Human Augmented Mapping** (HAM) that corresponds to a specific approach to support a robot in acquiring representations of the environment, in order to associate symbols to objects and locations perceived by the robot. These representations enable the robot to actually execute commands like "*go to my bedroom*", without being tele-operated by the user or requiring him to specify a target position in terms of coordinates. This process provides a general framework that does not depend from the underlying platform, also improving the adoption of a map in different robots. Moreover, it enables an incremental construction of the representation, as well as its revision in accordance with the changes in the environment [6]. In the HAM process, dialogue is crucial for a natural interaction between the user and the robot, especially when some properties of the entities or the environment itself cannot be directly derived from the sensory apparatus (e.g., whether an object is fragile or not).

Finally, we considered the **Task Teaching** process, that involves the interaction between the user and the robot to teach complex commands, which can be composed by primitive actions. In this respect, dialogue can support the extension of previous approaches by enabling the robot to learn parametric commands, as well as exploiting the knowledge about tasks to simplify the learning process [8, 31].

Hereafter, we will discuss a possible dialogue-based framework for Human-Robot Interaction to support the above tasks.

---

[1]A video describing this example is available at https://goo.gl/bpXmln and the underlying system is presented in [30].

3.2 A Framework for Flexible Pragmatic Task-Based Dialogues

We propose the adoption of an approach that we consider, to some extent, to be *pragmatic*. In fact, the final aim of the dialogue is to fulfill the information required to accomplish a given task, regardless it is an activity required to the robot or a step in the overall interpretation/mapping process.

We will adhere to the theory of *Information State* [32] for the management of the dialogues between the user and the robot. Such theory contemplates **informational components** (i.e. description of the context shared by the participants), **formal representation** of the aforementioned components, **dialogue moves** that trigger the update of such information, the **update rules** to be applied and the **update strategy** that is supposed to trigger the proper update rule.

The proposed framework will thus rely on the above (general) definitions to allow an easy and cost-effective design of dialogic interactions, specialized for a targeted task. These ideas are reflected by our framework, that is sketched in Fig. 1 and described hereafter.

The first module to be invoked during the processing of a user's utterance is the *Dialogue Act Classifier*: it extracts the intent of the user, expressed as a subset of Dialogue Acts (DA) proposed in [33]. This module gets the transcription of spoken utterances from Automatic Speech Recognition (ASR). Once the intent of a sentence
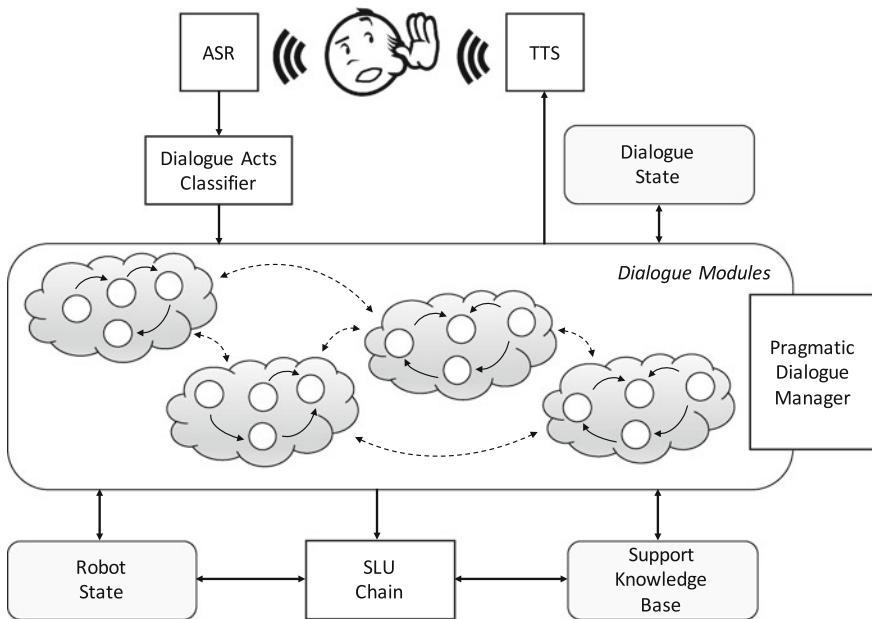

Fig. 1 The proposed framework for pragmatic task-based dialogue modeling

has been extracted (i.e. the user needs), the control of the dialogue is delivered to the *Pragmatic Dialogue Manager*, that controls the dialogue flow. It activates task-based representation structures (*Dialogue Modules*) that are in charge of fulfilling the missing information to accomplish the task. Such modules can be realized as (Partially Observable) Markov Decision Processes (POMDPs) or relying on Petri Net Plans (PNPs). The second solution allows to take into account the robot behavior and to harmonize the dialogue flow with the actions performed by the platform. Regardless of the implementation, possible interrupts are considered in the dialogue flow, to allow the user to control the overall dialogue and to facilitate timely reactions of the robotic platform to the user needs.

The status of the overall dialogue is traced by the Pragmatic Dialogue Manager by updating the *Dialogue State*, which stores aspects of the dialogue, such as the shared context and the parameters required by the robot to accomplish a given task. We decoupled such information with other aspects that are strictly related to the robot, namely the *Robot State* and the *Support Knowledge Base*. While the former collects physical and abstract aspects of the robot (e.g. manipulator availability, inability to perform a task, …), the latter maintains a structured representation of the environment, formalized through semantic maps and domain models. These resources are employed by the *Spoken Language Understanding Chain* (SLU Chain), that produces an interpretation of a user's utterance. The adopted SLU Chain [30] makes the interpretation process dependent also on the robot capabilities and the environmental settings, such as existence of entities referred in a user utterance, as well as spatial relations among them.

4   Use Cases

We identified several situations where such a framework can be used. Such scenarios are summarized hereafter and a more detailed use-case (related to the Human Augmented Mapping task) is reported at the end of this section.

Reasoning about the environment. In order to enable a semantic-aware navigation of the environment, the robot needs a structured representation of the world in which it operates. This representation is often built by relying on the interaction with the user, that instruct the robot for the operating environment. Often, this representation presents some mismatches with the real world, e.g. a book that is not into the semantic map but it is present the real environment. When the user asks to *Take the book*, the robot is supposed to start a dialogue to detect the book position, add it into the semantic map and complete the task.

Management of robot's self-awareness. Another situation where the dialogue is able to recover from undesired situations is when the robot is aware of its state (e.g. busy tray or manipulator, capability to perform some actions, …) and use a dialogic interaction to solve possible issues. In this case, when the robot detects a mismatch

between the user needs and its state, it should be able to leverage an interaction to solve potential hindrances.

Dealing with persisting ambiguities. In [13], a lexical grounding function has been proposed. Such function is able to link lexical references to real object, by observing linguistic aspects of the referring objects. We propose to use the dialogue to solve uncertainties about objects in the environment. A common scenario is when the user asks for an object and two entities of the same type are present. A proper interaction is supposed to solve such ambiguities.

Dialogue-based Human Augmented Mapping and Task Teaching. Each of these two tasks (Human Augmented Mapping and Task Teaching) required an ad-hoc interaction to accomplish the desiderata. In these cases, the dialogue can be employed for two purposes: (i) to acquire all the information needed in order to add a new entity to its semantic representation of the environment and (ii) to instruct the robot to perform new and unknown actions.

Interaction with user for re-training. The last scenario we are dealing with is when the sentence is misinterpreted. Assuming that our interpretation chain is based on Machine Learning techniques, correcting of a wrong interpretation represents a further step toward a system that is able to continuously learn by mistakes and improve its accuracy as interactions occurs. In this case, a suitable dialogue can lead the user to provide the correct interpretation of a sentence and this new observation can be employed to re-train the models the chain relies on.

4.1 Dialogue for Human Augmented Mapping

In Human Augmented Mapping, the user instructs the robot in constructing a structured representation of the environment. In this representation symbols are associated to the objects and locations in the environment. This operation is performed through interactions in Natural Language, where the robot learns the entities to be included in the semantic map. To this end, the sentences uttered by the user provide a mechanism to acquire the symbolic representation of the targeted object and location that populates the knowledge base.

When the robot is idle, the user can give commands to the robot. Once the user's intent and a structured representation of the sentence meaning are extracted by the SLU Chain, the robot attempts to ground each entity within the command. If an argument of the command denoting an object/location can not be grounded, the robot asks the user to help him in acquiring the position and features of the object/location. During this process, the user drives the robot with spoken commands such as *Turn right*, *Follow me* or *Go to the laboratory*. When the robot stands in front of the object/location that needs to be grounded, the user points to the object[2] and, once the

---

[2]This is currently done by a laser pen.

dot is recognized by a dot detection system, tells the robot the referring expression for the object, e.g., *Robot, this is the book*.

In the example below, the user refers to an object (a *book*) that is not present in the map. In other words, the robot does not know where the *book* is. Hence, the command *Bring me the book* can not be executed without the information about the existence of a book. The Human Augmented Mapping allows to complete such lack of knowledge and accomplish the task. To this end, once the request is given to the robot, the Dialogue Manager activates the dialogue module related to the Human Augmented Mapping.

User: Hi robot, I'm Andrea.
Robot: Hi Andrea. How can I help you?
User: Yes. Could you please bring me the book on the table?
Robot: I'm sorry, but I don't see any book. Could you help me?
User: Yes, sure. Follow me!
Robot: Okay, I'm following you!
                    [The robot starts following the user]
User: Okay, here we are.
                    [The robot stops following]
User: Now turn left.
                    [The robot turns left and the user points the book]
User: This is the book.
Robot: I suppose that I am able to grasp the book. Is it correct?
User: Yes, it is.
                    [The robot label the Graspable property of the book
                     as true]
Robot: Thank you, now I know what to bring you.
...

The Dialogue Manager is also in charge of handling interrupts of the dialogue flow when commands that are not exclusively related to the Human Augmented Mapping, such as *Follow me* or *Turn left*, are provided. Finally, once the user points the object and assigns a lexical reference to it (i.e., *book*), the Support Knowledge Base is updated with this information (i.e., position and affordances), also enabling the interpretation of future commands.

Within the same scenario, the next example shows how a dialogue can be used to solve ambiguous spatial references in the environment. In fact, we consider the same request of the previous example to a robot that does not have any reference of the book in the map, but it has references of two different tables:

User: Hi robot.
Robot: Hi. Can I help you?
User: Yes. Bring me the book on the table.
Robot: I'm sorry, but I don't see any book. Could you help me?
User: Yes. The book is on the table.
Robot: I see two tables in my map. Which one are you referring to?
User: The table of the laboratory.
Robot: Okay. I'm going to take the book in the laboratory. Please, wait.
User: Okay, thank you.
...

Notice that the Dialogue Manager deals with persisting ambiguities of the interpretation. In fact, the sentence *The book is on the table* is inherently ambiguous, whenever two tables co-exist in the same environment.

In conclusion, the above examples show different pragmatic interactions with different objectives. At the very beginning, the greeting dialogue is always activated whenever the user introduces himself. Then, once the intent of the user is understood, the request is delivered to the specific dialogue module, until the request is accomplished. Additionally, during the interactions the Dialogue Manager activates the flows to (i) fill missing knowledge, e.g., the robot does not know where is the book, (ii) drive it through specific commands, e.g., following and turning, (iii) characterize the entity, e.g., affordances, and (iv) resolve persisting ambiguities, e.g. the robot attempts to get the referred table. These features are essential to enable a natural interaction when teaching robots as in a Symbiotic Autonomy approach.

5 Conclusion

In this article, we propose a pragmatic framework aiming at effectively modeling dialogues within robotics platforms. The proposed approach aims at providing a natural interaction between a user and a robot by jointly exploiting (i) contextual information acquired by the robot, i.e., from the semantic map reflecting a semantically enriched representation of robot perceptions, (ii) knowledge related to the task to be accomplished, and (iii) other knowledge essential when dealing with robotic platforms, i.e., the robot state. Additionally, the framework allows to incrementally expand such resources, resulting in a more accurate and natural interaction with a robot that adapts itself to the user's profile.

The framework is based on the theory of Information State and the resulting architecture is decoupled in several task-based modules that are designed to support the robot in accomplishing the user's requests. The resulting architecture is thus biased towards the information required by the robot to determine the objectives of each interaction. The framework relies on a perceptually informed Spoken Language Understanding Chain to extract a structured representation of the meaning of user's utterances. In fact, such a chain exploits contextual information, e.g., existence of entities within the environment and spatial relations among them, to provide unambiguous interpretations and groundings that indeed depend on the environment where the interactions arise, as discussed in [17].

In order to support the potential contribution of the proposed approach, we identified some scenarios, in the context of Symbiotic Autonomy, that can benefit from the adoption of this framework, ranging from solving possible ambiguities of a command up to dialogue-based interactions for Human Augmented Mapping.

References

1. Rosenthal, S., Biswas, J., Veloso, M.: An effective personal mobile robot agent through symbiotic human-robot interaction. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, vol. 1, pp. 915–922. International Foundation for Autonomous Agents and Multiagent Systems (2010)
2. Bohus, D., Horvitz, E., Kanda, T., Mutlu, B., Raux, A.: Introduction to the special issue on dialog with robots. AI Mag. **32**(4), 15–16 (2011)
3. Winograd, T.: Procedures as a representation for data in a computer program for understanding natural language. Cogn. Psychol. **3**(1), 1–191 (1972)
4. Zuo, X., Iwahashi, N., Funakoshi, K., Nakano, M., Taguchi, R., Matsuda, S., Sugiura, K., Oka, N.: Detecting robot-directed speech by situated understanding in physical interaction. Trans. Jpn. Soc. Artif. Intell. **25**(6), 670–682 (2010)
5. Kruijff, G., Zender, H., Jensfelt, P., Christensen, H.: Situated dialogue and understanding spatial organization: knowing what is where and what you can do there. In: Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006, ROMAN 2006, pp. 328–333 (2006)
6. Bastianelli, E., Bloisi, D., Capobianco, R., Cossu, F., Gemignani, G., Iocchi, L., Nardi, D.: On-line semantic mapping. In: Proceedings of the 16th International Conference on Advanced Robotics (ICAR). pp. 1–6 (2013)
7. Rybski, P.E., Yoon, K., Stolarz, J., Veloso, M.M.: Interactive robot task training through dialog and demonstration. In: Proceedings of the ACM/IEEE International Conference on Human-robot Interaction, pp. 49–56. HRI '07, ACM, New York, NY, USA (2007)
8. Gemignani, G., Bastianelli, E., Nardi, D.: Teaching robots parametrized executable plans through spoken interaction. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, pp. 851–859. AAMAS '15, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2015)
9. Gemignani, G., Klee, S.D., Veloso, M., Nardi, D.: On task recognition and generalization in long-term robot teaching. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems. pp. 1879–1880. AAMAS '15, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2015)
10. Lemaignan, S., Alami, R.: "talking to my robot": From knowledge grounding to dialogue processing. In: Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction. pp. 409–410. HRI '13, IEEE Press, Piscataway, NJ, USA (2013)
11. Kollar, T., Tellex, S., Roy, D., Roy, N.: Toward understanding natural language directions. In: Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction, pp. 259–266. HRI '10, IEEE Press, Piscataway, NJ, USA (2010)
12. MacMahon, M., Stankiewicz, B., Kuipers, B.: Walk the talk: Connecting language, knowledge, and action in route instructions. In: Proceedings of the 21st National Conference on Artificial Intelligence, vol. 2, pp. 1475–1482. AAAI'06, AAAI Press (2006)
13. Bastianelli, E., Castellucci, G., Croce, D., Basili, R., Nardi, D.: Effective and robust natural language understanding for human-robot interaction. In: Proceedings of 21st European Conference on Artificial Intelligence, pp. 57–62. IOS Press (2014)
14. Bastianelli, E., Croce, D., Basili, R., Nardi, D.: Using semantic maps for robust natural language interaction with robots. In: Sixteenth Annual Conference of the International Speech Communication Association (2015)
15. Vanzo, A., Croce, D., Bastianelli, E., Basili, R., Nardi, D.: Robust spoken language understanding for house service robots. In: Proceedings of the 17th International Conference on Intelligent Text Processing and Computational Linguistics CICLing 2016. Konya, Turkey, April 3–9, 2016
16. Bastianelli, E., Croce, D., Basili, R., Nardi, D.: Using semantic models for robust natural language human robot interaction. In: Gavanelli, M., Lamma, E., Riguzzi, F. (eds.) AI*IA 2015 Advances in Artificial Intelligence: XIVth International Conference of the Italian Association for Artificial Intelligence, pp. 343–356. Springer International Publishing (2015)

17. Bastianelli, E., Croce, D., Vanzo, A., Basili, R., Nardi, D.: A discriminative approach to grounded spoken language understanding in interactive robotics. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016. New York, USA, July 9–15, 2016
18. Chernova, S., DePalma, N., Breazeal, C.: Crowdsourcing real world human-robot dialog and teamwork through online multiplayer games. AI Mag. **32**(4), 100–111 (2011)
19. Kollar, T., Perera, V., Nardi, D., Veloso, M.M.: Learning environmental knowledge from task-based human-robot dialog. In: ICRA, pp. 4304–4309. IEEE (2013)
20. Fang, R., Liu, C., She, L., Chai, J.Y.: Towards situated dialogue: revisiting referring expression generation. In: EMNLP, pp. 392–402 (2013)
21. Fang, R., Doering, M., Chai, J.Y.: Collaborative models for referring expression generation in situated dialogue. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 1544–1550 (2014)
22. Stoia, L., Shockley, D.M., Byron, D.K., Fosler-Lussier, E.: Noun phrase generation for situated dialogs. In: Proceedings of the Fourth International Natural Language Generation Conference, pp. 81–88. INLG '06, Association for Computational Linguistics, Stroudsburg, PA, USA (2006)
23. Garoufi, K., Koller, A.: Automated planning for situated natural language generation. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 1573–1582. ACL '10, Association for Computational Linguistics, Stroudsburg, PA, USA (2010)
24. Qu, S., Chai, J.Y.: Context-based word acquisition for situated dialogue in a virtual world. J. Artif. Intell. Res. 247–277 (2010)
25. Peltason, J., Wrede, B.: The curious robot as a case-study for comparing dialog systems. AI Mag. **32**(4), 85–99 (2011)
26. Marge, M., Rudnicky, A.I.: Miscommunication recovery in physically situated dialogue. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, p. 22 (2015)
27. Scheutz, M., Cantrell, R., Schermerhorn, P.W.: Toward humanlike task-based dialogue processing for human robot interaction. AI Mag. **32**(4), 77–84 (2011)
28. Brenner, M., Kruijff-Korbayová, I.: A continual multiagent planning approach to situated dialogue. In: Proceedings of the Semantics and Pragmatics of Dialogue (LONDIAL), p. 61 (2008)
29. Ross, R.J., Bateman, J.: Agency & information state in situated dialogues: analysis & computational modelling. In: Proceedings of the DiaHolmia, p. 113 (2009)
30. Bastianelli, E., Croce, D., Vanzo, A., Basili, R., Nardi, D.: Perceptually informed spoken language understanding for service robotics. In: Proceedings of the IJCAI2016 Workshop on Autonomous Mobile Service Robots. To appear (2016)
31. Klee, S.D., Gemignani, G., Nardi, D., Veloso, M.: Graph-based task libraries for robots: generalization and autocompletion. In: Proceedings of the 14th Italian Conference on Artificial Intelligence (2015)
32. Traum, D., Larsson, S.: The information state approach to dialogue management. In: Current and New Directions in Discourse and Dialogue, Text, Speech and Language Technology, vol. 22, pp. 325–353. Springer, Netherlands (2003)
33. Stolcke, A., Coccaro, N., Bates, R., Taylor, P., Van Ess-Dykema, C., Ries, K., Shriberg, E., Jurafsky, D., Martin, R., Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech. Comput. Linguist. **26**(3), 339–373 (2000)