A Consultant Framework for Natural Language Processing in Integrated Robot Architectures

Tom Williams

***Abstract*—One of the goals of the field of human-robot interaction is to enable robots to interact through *natural language*. This is a particularly challenging problem due to the uncertain and open nature of most application domains. In this paper, we summarize our recent work in developing natural language understanding and generation algorithms. These algorithms are specifically designed to handle the uncertain and open-worlds in which interactive robots must operate, and use a *Consultant Framework* specifically designed to account for the realities of integrated robot architectures.**

***Index Terms*—human-robot interaction; natural language understanding; natural language generation; natural language pragmatics; integrated robot architectures**

I. Introduction

ENGAGING in task-based natural language interactions in realistic situated environments is incredibly challenging [1]. This is especially true for robotic agents for three reasons. First, their knowledge is woefully *incomplete*, both physically (only having knowledge of a small number of objects, people, locations, and so forth) and socially (only having knowledge of a small number of social norms). Second, their knowledge is highly *uncertain* due perceptual and cognitive limitations. Finally, natural language is highly *ambiguous*. As such, language-enabled robotic architectures must be designed to handle uncertainty, ignorance, and ambiguity at each stage of the natural language pipeline.

In this article, we will summarize research performed in the Human-Robot Interaction laboratory at Tufts University in service of this goal[1]. Specifically, we will discuss research intending to account for uncertainty, ignorance, and ambiguity in the *referential* and *pragmatic* components of our robot architecture, DIARC [3], as implemented in the Agent Development Environment (ADE) [4][2]. We will begin by describing our *mnemonic architecture*: a *Consultant Framework* designed to facilitate domain independent memory retrieval in uncertain, open and ambiguous worlds. We will then describe our *linguistic architecture*: the language-processing components of our robot architecture which benefit from those mnemonic design choices.

Tom Williams was with the Department of Computer Science at Tufts University, Medford MA 02145 USA. E-mail: williams@cs.tufts.edu (See also: http://inside.mines.edu/~twilliams).

[1]Specifically, we summarize work that contributed to the author's doctoral dissertation [2]

[2]While our laboratory has also examined these topics at other stages of the natural language pipeline [5], that work is beyond the scope of this review.

II. Mnemonic Architecture

A. Distributed Heterogeneous Knowledge Bases

In integrated robot architectures (e.g., [3], [6]), information may be distributed across a variety of architectural components. Information about objects may be stored in one location, information about locations in another, and information about people and social relationships in yet another. Furthermore, each of these stores of knowledge may use a very different representational framework. The set of architectural components capable of providing information about entities that may be referenced in dialogue can thus be viewed as a set of *distributed, heterogeneous knowledge bases* [7]. Our mnemonic architecture makes the following assumptions [2] about such *DHKBs*: (1) Each DHKB has knowledge of some set of *entities*; (2) A subset of knowledge regarding each entity can be *accessed* through introspection and described using positive arity predicate symbols; (3) Any knowledge that can be accessed in this way can also be *assessed* as to strength of belief; and (4) Each sort of knowledge that can be accessed, assessed, and described can also be *imagined* to hold for a given entity.

B. Consultants

These assumptions are exploited using a set of *Consultants*. Each architectural Consultant provides domain-independent access to a particular DHKB, allowing other components to access, assess, and imagine entities without needing to know anything about how such entities are represented (see also Fig. 1). To facilitate this, each Consultant provides four capabilities: (1) providing a set of atomic entities assessable through introspection; (2) advertising a list of predicates that can be assessed with respect to such entities, listed according to a descending *preference ordering*; (3) assessing the extent to which it is believed such properties apply to such entities; and (4) imagining new entities and asserting knowledge regarding them.

III. Linguistic Architecture

Now that we have described DIARC's mnemonic architecture, we are ready to describe the linguistic architecture that leverages it (see also Fig. 2). We will begin by discussing natural language understanding components (reference resolution and pragmatic understanding), and then discuss natural language generation components (pragmatic generation and referring expression generation).

Throughout this section, we will use the example utterance "I need the medkit that is on the shelf in the breakroom", and a

mnemonic architecture using three DKHBs (*Mapping, Vision, Social*) and three associated consultants (*locs,objs,ppl*).

A. Reference Resolution

The first architectural component we will discuss is our *Reference Resolution* Component, whose job is to ascertain the identities of any entities referenced through natural language. For example, upon receiving the semantic representation for "I need the medkit that is on the shelf in the breakroom" $(Statement(speaker, self, need(self, X), \{on(X, Y), medkit(X), shelf(Y), breakroom(Z), in(Y, Z)\})$, it is up to the Reference Resolution Component to determine what entities should be associated with variables X, Y and Z. This problem can be broken down into three levels: closed-world reference resolution, open-world reference resolution, and anaphora resolution. In the following subsections we will discuss algorithms for solving each of these increasingly larger problems.

1) Closed-World Reference Resolution: Closed-World Reference Resolution is the basic problem of finding the optimal mapping from *references* to *known entities*. Under a simplifying assumption of inter-constraint independence, Closed-World Reference Resolution can be modeled using the following equation, where $\Lambda = \{\lambda_0, \dots \lambda_n\}$ is a set of semantic constraints, $\tilde{\Gamma} = \{\Gamma_0, \dots, \Gamma_n\}$ is the set of possible bindings to the variables contained in those constraints, and $\Phi = \phi_0, \dots, \phi_n$ is a set of satisfaction variables for which each ϕ_i is True iff formula λ_i holds under a given binding:

$$\Gamma\star = \underset{\Gamma \in \tilde{\Gamma}}{\operatorname{argmax}} \prod_{i=0}^{|\Lambda|} P(\phi_i \mid \Gamma, \lambda_i)$$

That is, the optimal set of bindings $\Gamma\star$ is that which maximizes the joint probability of a set of satisfaction variables Φ being satisfied under that binding and the set of provided semantic constraints Λ (under a simplifying assumption of independence between constraints). This model is algorithmically realized using the *DIST-CoWER* algorithm [2], [7], [8], which searches through the space of possible variable-entity assignments, pruning branches whose incrementally computed probability falls below a given threshold. For the example sentence, if the robot knows of a medkit on a shelf in a breakroom, it might return a set of bindings such as $\{X \rightarrow objs_4, Y \rightarrow objs_6, Z \rightarrow locs_9\}$, along with an associated probability value.

2) Open-World Reference Resolution: DIST-CoWER facilitates reference resolution under uncertainty, but presumes that all entities that could be referenced are known a priori: an assumption which is unwarranted in realistic human-robot interaction scenarios. To continue the previous example, if a robot knows of a breakroom but does not know of a shelf in that breakroom, DIST-CoWER will fail to return any bindings for "the medkit that is on the shelf in the breakroom". Ideally, however, the robot would be able to both resolve the portions of the utterance that it *does* know of a priori, and learn in one shot about other, previously unknown entities referenced in the utterance. We model this problem, which we call *Open-World Reference Resolution*, using the following equation, where Λ is a set of constraints, Λ^V is an ordering of the variables involved in those constraints, $\Gamma^i\star$ is the optimal solution provided by DIST-CoWER given the constraints involving only the last $|\Lambda^V| - i$ variables in Λ^V, $\Gamma^i\star^P$ is the probability associated with this solution, and $complete$ is a function which creates new representation to associate with any variables appearing in Λ^V but not in optimal open-world solution $\Gamma^i\star$:

$$\Gamma\star = complete\left(\underset{\Gamma^i\star \in \{\Gamma^0\star, \dots, \Gamma^{|\Lambda^V|}\star\}}{\operatorname{argmax}} \begin{cases} i, & \text{if } \Gamma^i\star^P > \tau \\ 0, & \text{otherwise} \end{cases}\right).$$

That is, the optimal set of bindings $\Gamma\star$ is the first sufficiently probable set of bindings returned from a series of calls to DIST-CoWER made using different subsets of the full predicate set Λ. This model is algorithmically realized using the *DIST-POWER* algorithm [2], [7], [8], which (1) finds a variable ordering over the variables used in predicate list Λ based on linguistic factors such as prepositional attachment; (2) successively removing variables from this list until calling *DIST-CoWER* on only the predicates involving the remaining variables returns a sufficiently probable solution; (3) for each variable removed in this way, instructing the appropriate consultant to create a mental representation for a new entity; (4) instructing the appropriate Consultants to make any representations created in this way to be consistent with all related predicates in Λ; (5) returning a unified set of bindings from the set of variables used in Λ to known and/or newly created entity representations.

For example, if in the example sentence the robot does not know of a shelf in a breakroom, it might return the set of bindings $\{X \rightarrow objs_{44}, Y \rightarrow objs_{45}, Z \rightarrow locs_9\}$, where $objs_{44}$ and $objs_{45}$ are references to newly created representations for hypothesized entities, and $locs_9$ is a reference to a previously existing representation for a (grounded or hypothetical) entity.

3) Anaphora Resolution: The algorithms discussed in the previous sections allow our robots to resolve references in uncertain and open worlds using *definite noun phrases* (i.e., "the-*N*" phrases). Humans, however, have a tendency to use a much wider variety of referring forms, (e.g., "a-*N*", "this", "it"). In order to handle this multitude of forms, we embed *DIST-POWER* within a larger algorithm called *GH-POWER* [9] for its use of the *Givenness Hierarchy* theory of reference. The Givenness Hierarchy divides referential forms into six groups, each of which is associated with a different tier of a hierarchy of six nested cognitive statuses. For example, when one uses "it", the Givenness Hierarchy suggests that the speaker believes their target referent to be at least *in focus* for the listener; when one uses "this", the Givenness Hierarchy suggests the speaker believes their target referent to be at least *activated*, that is, in *short-term memory* for the listener (and possibly also *in focus* because of the nested nature of the six tiers); and so forth. Using this theory, when a particular referential form is used, we infer the set of possible statuses the speaker may believe their target referent to have, and from this infer a sequence of *mnemonic actions* to take: creating a new representation of a referent or searching for an existing one in a particular data structure. For example, when a definite noun phrase is used, *GH-POWER* will search through the set

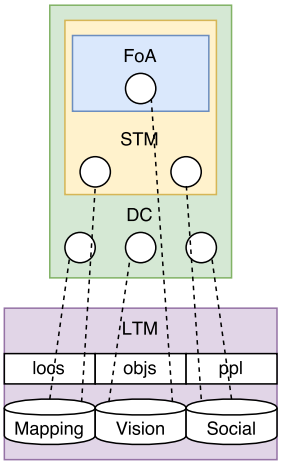

Fig. 1. **Memory Model:** The Focus of Attention, Short Term Memory, Discourse Context are hierarchically nested, and contain references to entities stored in the Distributed, Heterogeneous Knowledge Bases which comprise long-term memory (*Mapping, Vision, Social*), access to which is enabled and controlled using a set of Consultants (*locs, objs, ppl*).

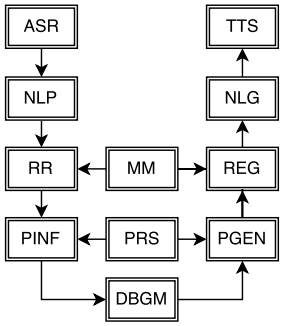

Fig. 2. **Architectural Diagram:** Information flows through the following components: Automatic Speech Recognition (ASR), Natural Language Processing (NLP), Reference Resolution (RR), Pragmatic Understanding (PUND), the Dialogue Belief and Goal Manager (DBGM), Pragmatic Generation (PGEN), Referring Expression Generation (REG), Natural Language Generation (NLG), Text to Speech (TTS). Utilized by these components are the consultant framework and associated Memory Model (MM) and the Pragmatic Rule Set (PRS)

of activated entities, the focus of attention, the set of familiar entities, and, if none of these steps are successful, will use *DIST-POWER* to search through all of the robot's long term memory, and hypothesizing a new representation if that search fails. Fig. 1 shows the set of Givenness Hierarchy-theoretic data structures we use (Focus of Attention, Short Term Memory, Discourse Context), their hierarchical relationship with each other, and how the representations within each other can be viewed as references to entities stored in the DHKBs which comprise long-term memory (*Mapping, Vision, Social*), access to which is enabled and controlled using a set of Consultants (*locs, objs, ppl*).

For example, in the example sentence, all three entities are described using "the"; as such, *GH-POWER* will use the sequence of mnemonic actions Search STM, Search FoA, Search DC, Search LTM, Hypothesize, where these last two steps are achieved (if necessary) by performing a *DIST-POWER* query. If sufficiently probable candidate referents can be found in the upper level GH-theoretic data structures, this may not be necessary.

4) Discussion: To summarize thus far, given the unbound semantic interpretation of an incoming utterance, our referential components will first identify sequences of *mnemonic* actions to take to resolve the references found in that utterance; those sequences will then be used to initiate searches through various data structures for entities that, according to the appropriate Consultants, are likely to satisfy the predicates that comprise the semantic interpretation; in the worst case this will require the use of *DIST-POWER* to search all of long-term memory. Once the reference resolution process is completed, the end result is a set of hypotheses, each of which associate the variables found in the incoming semantic interpretation with a different set of entities, and each of which has a particular likelihood. As described in [10], these hypotheses are then used to create a set of *bound utterances*, which are combined into a single Dempster-Shafer theoretic *body of evidence* which is passed to our *pragmatic reasoning* component.

Our approach significantly differs from most other recent language understanding work in robotics. Most other work has focused on tackling the full *language grounding* problem of mapping from natural language to continuous perceptual representations [11]–[20]. In contrast, we separate this problem into two halves: *reference resolution*, in which natural language is mapped to discrete symbols representing unique entities, and *symbol grounding*, in which those symbols are associated with continuous perceptual representations, and focus on the development of reference resolution algorithms. This division has allowed us to develop a framework for resolving references both to grounded, observed entities, as well as to heretofore unknown or even hypothetical and imaginary entities, thus providing us the means to tackle *open-world* reference resolution.

B. Pragmatic Reasoning

In this section we will discuss the Pragmatic Reasoning components of our architecture: Pragmatic Understanding and its counterpart, Pragmatic Generation.

1) Pragmatic Understanding: The pragmatic reasoning component's primary purpose is to infer the intentions behind incoming utterances. Specifically, this component attempts to infer the intentions behind conventionally indirect utterance forms, also known as *Indirect Speech Acts*, which recent work has shown to be prevalent throughout human-robot dialogue [21]. Provided with the Dempster-Shafer theoretic set of candidate utterances Θ_u produced by *GH-POWER*, a Dempster-Shafer theoretic set of contextual information Θ_c and a set of Dempster-Shafer theoretic rules of the form $u \wedge c \Rightarrow i$, the pragmatic reasoning component produces a Dempster-Shafer theoretic set of intentions Θ_i by computing $m_i(\cdot) = \left((m_u \otimes m_c) \odot m_{uc \to i}\right)(\cdot)$, where $\otimes$ is a Dempster-Shafer theoretic *And* operator, and $\odot$ is a Dempster-Shafer theoretic *Modus Ponens* operator [22].

For example, when processing the example sentence, a robot might have a pragmatic rule that says (with some probability between, say, 0.7 and 0.8) that when someone says to someone they believe to be their subordinate that they need something, they likely want their subordinate to bring them that something:

$$Stmt(A, B, need(A, C)) \wedge bel(A, subordinate(B, A)) \xrightarrow{[0.7,0.8]} want(A, goal(B, bring(B, C, A))).$$

Because the example utterance matches this rule's utterance form, the uncertainty interval reflecting the confidence that that utterance was what was actually said is combined with the uncertainty interval reflecting the robot's confidence that the speaker believes the robot to be their subordinate, as well as with the uncertainty interval associated with the rule itself,

producing an uncertainty interval reflecting how confident the robot is that the speaker wants it to bring them the described object.

Each candidate intention I inferred in this way is thus augmented with a Dempster-Shafer theoretic interval $[\alpha, \beta]$ within which the probability that I is true can be said to lie. If an interval is determined to reflect sufficient uncertainty by Nunez' uncertainty measure [23]

$$\lambda = 1 + \frac{\beta}{1+\beta-\alpha} log_2 \frac{\beta}{1+\beta-\alpha} + \frac{1-\alpha}{1+\beta-\alpha} log_2 \frac{1-\alpha}{1+\beta-\alpha},$$

the robot generates it's own intention – an intention to know whether or not it should actually infer that candidate intention. This allows the robot to identify sources of both pragmatic and referential uncertainty and ignorance. If such an intention is generated, it will satisfied by generating a *clarification request* [10]. This highlights the other capability of the pragmatic reasoning component: to perform *pragmatic generation*.

2) Pragmatic Generation: Due to our use of a Dempster-Shafer-theoretic approach, the same rules used to infer the intentions behind utterances (pragmatic understanding) can also be used to abduce utterances that can be used to communicate intentions (pragmatic generation). Pragmatic generation is performed using the same set of Dempster-Shafer theoretic rules and logical operators used during pragmatic understanding [22], with the addition of one pre-processing step and one post-processing step.

Before pragmatic generation is performed, it is determined whether the robot is generating a clarification request, and if so, whether there are more than two choices must be arbitrated between [10]. If so (e.g., if, in the example, the robot determines it knows of two medkits on a shelf in a breakroom), those choices are unified into a single predicate which will allow a generic WH-question (e.g., "*Which* medkit would you like?") rather than a many-item YN-question (e.g., "Would you like the red medkit or the blue medkit or the white medkit or the green medkit?"), which are only when there are two or fewer options. Note, however, that at this stage of processing, only the bare utterance form has been generated (e.g., QuestionYN(self,speaker, or(would(speaker,like(speaker,obj_1)), would(speaker,like(speaker,obj_2))))), and not the properties used to describe each referenced entity.

After pragmatic generation has yielded a set of utterance forms which could be communicated, each is passed *forwards* through the pragmatic reasoning module, in order to simulate the utterance understanding process. This creates a set of intentions the robot may believe its interlocutor will infer if it chooses to use that particular utterance form. This allows the robot to detect unintended side-effects of different candidate utterance forms so that it can choose the best possible utterance to communicate its intentions. The best candidate utterance form is then selected and sent to our *Referring Expression Generation* module.

3) Discussion: Our work on pragmatic understanding directly builds off of previous work from Briggs et al. [24]. Like that work, our own understands utterances based on its beliefs *about the speaker's beliefs*; but we improve on that work by handling uncertainty (and ignorance) and allowing for adaptation: capabilities also largely lacking from previous computational approaches to ISA understanding (e.g., [25]–[27]).

Our techniques for generating clarification requests compare favorably to previous work due to our accounting for human preferences (cf. [28], [29]) and our ability to handle uncertainty (cf. [12]). Similarly, while there has been some previous work on generating indirect language (e.g., [24], [30]), we believe that our work is the first to enable robots' generation of conventionalized indirect speech acts under uncertainty.

C. Referring Expression Generation

Once a robot has chosen an utterance form to communicate, it must decide what properties to use to describe the things it wishes to communicate about. This is a problem known as Referring Expression Generation. To solve this problem, we use *DIST-PIA* [31], a version of the classic *Incremental Algorithm* [32], modified to use our consultant framework. When crafting a referring expression for a given entity, our algorithm proceeds through the ordered list of properties provided by the consultant responsible for that entity: each property is added to the list of properties to be used in the description if it is sufficiently probable that it applies to the target referent, and if it is not sufficiently probable that it applies to one or more *distractors*, thus allowing those distractors to be ruled out. This algorithm improves on the classic Incremental Algorithm in its use of our Mnemonic Architecture (to enable use in integrated robot architectures) and in its ability to operate under uncertainty.

For example, suppose the *objs* consultant advertises that it can handle the following properties: $\{shelf(X - objs), medkit(X - objs), green(X - objs), red(X - objs), on(X - objs, Y - objs), in(X - objs, Y - locs)\}$, and that the *DIST-PIA* algorithm is used to generate a description of $objs_4$. Suppose $objs_4$ is not likely to be a shelf – that property will be ignored. Suppose $objs_4$ is likely to be a medkit, as are two other objects – $medkit(objs_4)$ will be added to the set of properties to use. Suppose $objs_4$ is not likely to be green – that property will be ignored. Suppose $objs_4$ is likely to be red, and that neither of the two *distractor* medkits are likely to be so – $red(objs_4)$ will be added to the set of properties to be used, and since all distractors have been eliminated, the set $\{medkit(objs_4), red(objs_4)\}$ will be returned, allowing a description such as "the red medkit" to be generated.

Of equal contribution in the work that presented this algorithm [31] is our development of a novel evaluation framework that solicits certainty estimates from humans in order to craft probability distributions that can then be used by uncertainty-handling REG algorithms. This allows such algorithms to be evaluated with respect to both other algorithms and human beings, without committing to any particular set of visual classifiers (cf. [33]–[35]).

IV. Conclusion

In this article, we have described a natural language understanding and generation pipeline designed specifically for use

in integrated robot architectures and for operation in uncertain and open worlds. It is our hope that the general frameworks presented in this work will allow researchers to more easily integrate together disparate approaches – and that this work will draw researchers' attention to the under-studied area of open-world language processing. For a complete treatment of the work described in this paper, we direct the interested reader to the authors' recent dissertation, which describes the work cited herein in much more detail [2].

ACKNOWLEDGMENT

This work was funded in part by ONR grants #N00014-11-1-0289, #N00014-11- 1-0493, #N00014-10-1-0140 #N00014-14-1-0144, #N00014-14-1-0149, #N00014-14-1-0751, and NSF grants #1111323, #1038257.

REFERENCES

[1] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, 2015.

[2] T. Williams, "Situated natural language interaction in uncertain and open worlds," Ph.D. dissertation, Tufts University, 2017.

[3] P. W. Schermerhorn, J. F. Kramer, C. Middendorff, and M. Scheutz, "DIARC: A testbed for natural human-robot interaction." in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2006, pp. 1972–1973.

[4] M. Scheutz, G. Briggs, R. Cantrell, E. Krause, T. Williams, and R. Veale, "Novel mechanisms for natural human-robot interactions in the diarc architecture," in *Proceedings of AAAI Workshop on Intelligent Robotic Systems*, 2013.

[5] M. Scheutz, E. Krause, B. Oosterveld, T. Frasca, and R. Platt, "Spoken instruction-based one-shot object and action learning in a cognitive robotic architecture," in *Proceedings of the Sixteenth International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2017.

[6] M. Quigley, J. Faust, T. Foote, and J. Leibs, "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.

[7] T. Williams and M. Scheutz, "A framework for resolving open-world referential expressions in distributed heterogeneous knowledge bases," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[8] ——, "POWER: A domain-independent algorithm for probabilistic, open-world entity resolution," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[9] T. Williams, S. Acharya, S. Schreitter, and M. Scheutz, "Situated open world reference resolution for human-robot dialogue," in *Proceedings of the Eleventh ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016.

[10] T. Williams and M. Scheutz, "Resolution of referential ambiguity in human-robot dialogue using dempster-shafer theoretic pragmatics," in *Proceedings of Robotics: Science and Systems (RSS)*, 2017.

[11] P. Gorniak and D. Roy, "Grounded semantic composition for visual scenes," *Journal of Artificial Intelligence Research*, vol. 21, pp. 429–470, 2004.

[12] G.-J. M. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, and N. Hawes, "Incremental, multi-level processing for comprehending situated dialogue in human-robot interaction," in *Symposium on Language and Robots*, 2007.

[13] S. Lemaignan, R. Ros, R. Alami, and M. Beetz, "What are you talking about? grounding dialogue in a perspective-aware robotic architecture," in *The Eighteenth IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2011, pp. 107–112.

[14] F. Meyer, "Grounding words to objects: A joint model for co-reference and entity resolution using markov logic for robot instruction processing," Ph.D. dissertation, TUHH.

[15] J. Y. Chai, L. She, R. Fang, S. Ottarson, C. Littley, C. Liu, and K. Hanson, "Collaborative effort towards common ground in situated human-robot dialogue," in *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction (HRI)*, 2014, pp. 33–40.

[16] J. Fasola and M. J. Matarić, "Interpreting instruction sequences in spatial language discourse with pragmatics towards natural human-robot interaction," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 2720–2727.

[17] C. Kennington and D. Schlangen, "A simple generative model of incremental reference resolution for situated dialogue," *Computer Speech & Language*, vol. 41, pp. 43–67, 2017.

[18] S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, A. G. Banerjee, S. Teller, and N. Roy, "Approaching the symbol grounding problem with probabilistic graphical models," *AI Magazine*, 2011.

[19] I. Chung, O. Propp, M. R. Walter, and T. M. Howard, "On the performance of hierarchical distributed correspondence graphs for efficient symbol grounding of robot instructions," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 5247–5252.

[20] C. Matuszek, E. Herbst, L. Zettlemoyer, and D. Fox, "Learning to parse natural language commands to a robot control system," in *Proceedings of the Thirteenth International Symposium on Experimental Robotics (ISER)*, 2012.

[21] G. Briggs, T. Williams, and M. Scheutz, "Enabling robots to understand indirect speech acts in task-based interactions," *Journal of Human-Robot Interaction*, 2017.

[22] T. Williams, G. Briggs, B. Oosterveld, and M. Scheutz, "Going beyond command-based instructions: Extending robotic natural language interaction capabilities," in *Proceedings of 29th AAAI Conference on Artificial Intelligence*, 2015.

[23] R. C. Núñez, R. Dabarera, M. Scheutz, G. Briggs, O. Bueno, K. Premaratne, and M. N. Murthi, "DS-Based Uncertain Implication Rules for Inference and Fusion Applications," in *Sixteenth International Conference on Information Fusion*, July 2013.

[24] G. Briggs and M. Scheutz, "A hybrid architectural approach to understanding and appropriately generating indirect speech acts," in *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[25] D. J. Litman and J. F. Allen, "A plan recognition model for subdialogues in conversations," *Cognitive science*, vol. 11, no. 2, pp. 163–200, 1987.

[26] E. A. Hinkelman and J. F. Allen, "Two constraints on speech act ambiguity," in *Proceedings of the Twenty-Seventh annual meeting of the Association for Computational Linguistics (ACL)*, 1989, pp. 212–219.

[27] S. Wilske and G.-J. Kruijff, "Service robots dealing with indirect speech acts," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006, pp. 4698–4703.

[28] R. Deits, S. Tellex, T. Kollar, and N. Roy, "Clarifying commands with information-theoretic human-robot dialog," *Journal of Human-Robot Interaction*, 2013.

[29] S. Hemachandra, M. R. Walter, S. Tellex, and S. Teller, "Learning spatial-semantic representations from natural language descriptions and scene classifications," in *International Conference on Robotics and Automation (ICRA)*, 2014.

[30] R. A. Knepper, C. I. Mavrogiannis, J. Proft, and C. Liang, "Implicit communication in a joint action," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. ACM, 2017, pp. 283–292.

[31] T. Williams and M. Scheutz, "Referring expression generation under uncertainty: Algorithm and evaluation framework," in *Proceedings of the Tenth International Conference on Natural Language Generation (INLG)*, 2017.

[32] R. Dale and E. Reiter, "Computational interpretations of the gricean maxims in the generation of referring expressions," *Cognitive science*, vol. 19, no. 2, pp. 233–263, 1995.

[33] H. Horacek, "Generating referential descriptions under conditions of uncertainty," in *Proceedings of the Tenth European Workshop on Natural Language Generation (ENLG)*, 2005, pp. 58–67.

[34] A. Sadovnik, A. Gallagher, and T. Chen, "Not everybody's special: Using neighbors in referring expressions with uncertain attributes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 269–276.

[35] R. Fang, C. Liu, L. She, and J. Y. Chai, "Towards situated dialogue: Revisiting referring expression generation." in *Proceedings of the Conference on Empirical Methods for Natural Language Processing (EMNLP)*, 2013, pp. 392–402.