Resolution of Referential Ambiguity Using Dempster-Shafer Theoretic Pragmatics

Tom Williams and Matthias Scheutz
Human-Robot Interaction Laboratory
Tufts University, Medford, MA, USA
{williams,mscheutz}@cs.tufts.edu

Abstract

A major challenge for robots interacting with humans in realistic environments is handling robots' uncertainty with respect to the identities and properties of the people, places, and things found in their environments: a problem compounded when humans refer to these entities using *underspecified* language. In this paper, we present a framework for generating clarification requests in the face of both pragmatic and referential ambiguity, and show how we are able to handle several stages of this framework by integrating a Dempster-Shafer (DS)-theoretic pragmatic reasoning component with a probabilistic reference resolution component.

Introduction

Imagine a robot named Cindy operating in a disaster relief scenario. Cindy and her human teammate Bob have just left a second-floor room containing a small refrigerator, a sink, and two medical kits: one on a table and one on a counter. After driving ten meters down the hallway, Bob says "Go back to the kitchen and grab the medical kit." In order to understand this command, Cindy must (1) resolve *referential uncertainty* by deciding how probable it is that the previous room was a kitchen or not, and (2) resolve *referential ambiguity* by deciding whether that room or another kitchen was being referred to (as well as which of the two medical kits was being referred to). In order to resolve referential uncertainty and/or ambiguity, Cindy may need to ask for clarification as a human would (Tenbrink et al. 2010). In this scenario, for example, Cindy might say "Do you want me to retrieve the medical kit on the counter or the medical kit on the table?", or "Do you mean the room we were just in, or the kitchen on the first floor?"

In previous work, we showed how a Dempster-Shafer (DS)-theoretic pragmatic reasoning component could be used to generate clarification questions under *intentional* uncertainty and ignorance. For example, if a robot was told "The commander needs a medical kit", and was unsure of the social relationship between itself and the speaker, it might identify two alternate interpretations and ask "Do you want me to bring him one, or do you want to know where to find one?" In this work, we show how this pragmatic reasoning component can also be used to identify *referential* uncertainty and ambiguity, and generate clarification requests appropriately. Specifically, we present a clarification request generation framework, and demonstrate how, by integrating our pragmatic reasoning component with a probabilistic reference resolution component, we are able to handle several stages of this framework.

In the next section, we discuss previous work on situated clarification request generation. Next, we lay out a theoretical clarification request generation framework. We then describe the stages of this framework handled by components of the Distributed, Integrated, Affect, Reflection and Cognition (DIARC) architecture (Scheutz et al. 2007), and demonstrate those components in operation. Finally, we describe how remaining stages of the framework might be handled in the future.

Related Work

Clarification request generation has been a topic of considerable research in non-situated contexts (Purver, Ginzburg, and Healey 2003; Traum 1994), but has only recently become a topic of interest in the Human-Robot Interaction community (c.f. general question asking capabilities, e.g., Fong, Thorpe, and Baur (2001), Rosenthal, Veloso, and Dey (2012)). Several recent approaches have used information-theoretic techniques to determine the best random variable of interest to ask questions about, with questions either framed as yes/no questions (Deits et al. 2013; Hemachandra, Walter, and Teller 2014; Purver 2004) or specification requests (e.g., "What do the words X refer to?") (Tellex et al. 2013; Purver 2004) However, recent experimental evidence (Marge and Rudnicky 2015) suggests that people prefer robots to list multiple options rather than confirm a single referent with a yes/no question (c.f. Clark (1996)), even when asking a yes/no question would be more efficient (c.f. Hemachandra, Walter, and Teller (2014)).

Perhaps closest to the proposed approach is that presented by Kruijff, Brenner, and Hawes (2008). Those authors resolve ambiguity using a continual planning

approach that makes use of actions generating utterances that list multiple options, such as "Do you mean the blue or the red mug, Anne?"

As we will describe, we take a similar approach. However, instead of directly planning to receive disambiguating information through communicative actions, we simply identify points of ambiguity; DIARC's dialogue system then decides as part of its normal functioning to pose questions to resolve this ambiguity, and does so in a way which accounts for social context, uncertainty, and ignorance, none of which appear to be handled by Kruijff et al.. Moreover, because DIARC's dialogue system uses a single set of pragmatic rules for understanding and generation, the same rules that allow the robot to generate clarification requests also allow the robot to understand such requests.

There has also been much previous work in the area of natural language generation (NLG). Most broadly relevant perhaps are general NLG frameworks like that of Reiter, Dale, and Feng (2000)). Dale and Reiter identify six stages of NLG: content determination, document structuring, aggregation, lexical choice, referring expression generation, and realisation. We would argue, however, that NLG *for use in Human-Robot Interaction* warrants a framework that deviates from that used for more traditional NLG purposes. In HRI, NLG typically needs to happen at the utterance level rather than at the document level, which deemphasizes steps such as document structuring and aggregation. More fundamentally, NL is generated for different reasons in HRI than it is in other contexts: it is more likely in HRI than in other contexts that utterances must be generated to *solicit*, rather than *provide*, information, e.g., through *clarification requests*. In the next section we present an NLG framework specifically designed to facilitate clarification request generation in HRI contexts.

A Framework for Clarification Request Generation

We identify five stages necessary for successful clarification request generation, as shown in Fig. 1: (1) uncertainty identification, (2) decision to communicate, (3) utterance choice, (4) surface realization, and (5) speech synthesis. In this section we describe the actions necessary at each stage.

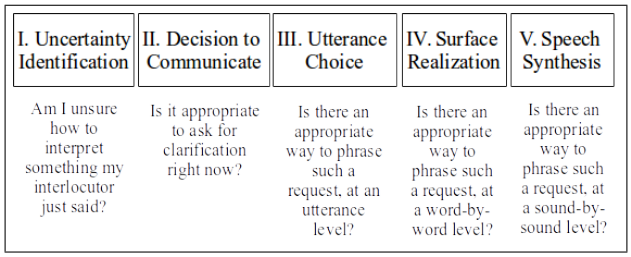

Figure 1: *Clarification Framework.*

Uncertainty Identification

Suppose that in our original example, Bob had asked Cindy "Can you grab the medkit?" During the stage of *uncertainty identification*, Cindy must determine if she is unsure how to interpret any part of this utterance. This may be uncertainty as to what entities are being *referenced*, e.g., *which* medkit Bob is referring to, or uncertainty as to the speaker's *intentions*, e.g., whether Bob wishes Cindy to bring him the medkit or whether he meant something else by the utterance. Furthermore, this uncertainty may take different forms (c.f. Stirling (2010)): the utterance may be *ambiguous* (e.g., if Cindy knows of multiple medkits) or the utterance may reveal *ignorance* (e.g., if Cindy knows of no medkits, or is unsure whether a particular object qualifies as a "medkit").

Decision to Communicate

If a robot has identified a point in need of clarification, it must decide whether it would be appropriate to actually ask for clarification. This decision will depend on a variety of factors: Is it permissible for the robot to ask for clarification? Is the robot's interlocutor likely to be able to provide clarification? Would obtaining clarification really be the highest utility action at the current time (compared to, e.g., exploration)? For example, if Cindy determines there are actually two medkits that Bob could be referring to, but while coming to this decision Bob has already engaged another teammate in conversation, it may be necessary for Cindy to wait until this conversation finishes before asking for clarification.

Utterance Choice

Once a robot has decided to request clarification on a particular point, it must determine what utterance form to use to communicate its request: depending on the relationship between the robot and its interlocutor, and the obligations of each party, certain utterance forms may be more or less appropriate (Brown 1987). For example, if Cindy if Bob's subordinate, it may be more appropriate to use an *indirect request* such as "Which medkit would you like?", whereas if Cindy is Bob's superior, it may be more appropriate to use a *direct request* such as "Tell me which medkit you would like."

Surface Realization

Once a robot chooses an utterance form to use, it must determine what words to use (Garoufi and Koller 2014). For example, if Cindy decides to use an utterance of the form "Would you like [medkit$_1$]", she must choose how to actually describe medkit$_1$, e.g., by referring to it as "the medkit in the kitchen" or perhaps as "the white medkit". If one medkit is in front of Cindy, it may be more appropriate to point to it and use a deictic expression such as "this medkit."

Speech Synthesis

Finally, once a robot determines what word to use, it must synthesize an appropriate sound pattern.

A DS-Pragmatic Approach

We have implemented the first three stages of the proposed framework as components of the DIARC architecture (Scheutz et al. 2007). In this section, we describe this implementation, and discuss how the fourth and fifth stages could be handled in future work.

Notation[1]

M A robot's *world model* of entities $\{m_0 \dots m_n\}$.

Λ A set of logical formulae $\lambda_0 \dots \lambda_n$, denoting (literal, direct) semantic *connotation* of an incoming utterance.

V A set of free variables found in Λ.

Γ A set of bindings from variables in V to entities in M, denoting the semantic *denotation* of an incoming utterance.

Φ A *satisfaction* variable which is *True* iff all formulae in Λ *hold* when bound using Γ.

Uncertainty Identification

Uncertainty may be identified at many stages along the natural language (NL) pipeline. For example, if a robot determines it is unsure what words were uttered by an interlocutor, it may immediately ask for clarification (Stoyanchev, Liu, and Hirschberg 2013). In this paper, however, we are specifically interested in *referential uncertainty and ambiguity*. In our implementation, the *Resolver Component* uses the *GH-POWER* algorithm in order to resolve referring expressions (Williams et al. 2016). As described in previous work, this algorithm uses a Givenness-Hierarchy (GH) theoretic approach (Gundel, Hedberg, and Zacharski 1993) to search models of cognitive structures (e.g., the Focus of Attention, Short-Term Memory, and Long-Term Memory) for the referents of referring expressions, including deictic and anaphoric expressions. Furthermore, the GH-POWER algorithm hypothesizes new representations for previously unknown referents when appropriate. Long-Term Memory queries are effected using the *DIST-POWER* algorithm, which allows us to distribute long term memories across heterogeneous knowledge bases stored on different machines (Williams and Scheutz 2016).

In addition to potentially hypothesizing new entities, POWER's ultimate result is the distribution $P(\Phi \mid \Gamma, \Lambda)$. That is, the probability of successful satisfaction conditioned on binding hypotheses

$$\{\Gamma_0 = \{\gamma_{0_0} \dots \gamma_{0_n}\} \dots, \Gamma_m = \{\gamma_{m_0} \dots \gamma_{m_n}\}\}$$

and semantic parse hypotheses

$$\{\Lambda_0 = \{\lambda_{0_0} \dots \lambda_{0_n}\} \dots, \Lambda_m = \{\lambda_{m_0} \dots \lambda_{m_n}\}\}.$$

Note that the parser we use only currently returns a single best parse; we use this notation to allow for the future possibility of multiple semantic interpretations.

[1] c.f. Tellex et al. (2011)

Next, consider the example utterance "Can you grab the medkit?". This may be parsed by Cindy into something like $QuestionYN(b, s, can(s, grab(s, X)))$ with additional semantic content $\Lambda_i = \{medkit(X)\}$, where b="bob" and s="self" (we will use these abbreviations throughout this section). If the robot is 70% sure that the object with identifier m_5 is a medkit, reference resolution will produce:

$$P(\Phi = True \mid \Gamma = \{X \rightarrow m_5\}, \Lambda = \{medkit(X)\}) = 0.7$$

The set of sufficiently probable referential hypotheses is then used to create a set of *bound utterances with supplemental semantics* (BUSSes) $\Psi = \{\psi_0 \dots \psi_n\}$ where each ψ_i is created by binding the free variables of the parsed utterance form (e.g., $QuestionYN(b, s, can(s, grab(s, X)))$) and supplemental semantics (e.g., $\{medkit(X)\}$) with variable bindings γ_i (e.g., $\{X \rightarrow m_5\}$), producing something like:

$$\{QuestionYN(b, s, can(s, grab(s, m_5))) \wedge medkit(m_5)\}.$$

While it would be possible to create a *distribution* over this set, where $P(\psi_i) = P(\Gamma_i, \Lambda_i \mid \Phi_i)$ using, e.g., Bayes' Rule, this would only be appropriate if the next component in the NL pipeline also used a Bayesian approach. In fact, the next component (i.e., the pragmatic reasoning component) actually uses a *Dempster-Shafer theoretic* approach (Williams et al. 2015).

Dempster-Shafer (DS) Theory is a generalization of the Bayesian uncertainty framework that allows for elegant reasoning about uncertainty and ignorance even when distributional information is not available (Shafer 1976). DS Theory is an attractive option for many robotics applications, where agents may need to learn about new entities and concepts from a small number of examples drawn from an unknown distribution.

Of course, not all of a robot architecture's components are likely to be DS-theoretic. For some components, distributional information may be readily available, encouraging the use of a Bayesian approach. To allow each architectural component to use the knowledge representation and uncertainty management approaches most conducive to its own operation, we must thus develop mechanisms that allow those components to integrate seamlessly. In the rest of this section, we will (1) briefly provide some preliminaries of DS Theory, (2) describe how it is used in our architecture, and (3) describe the technique we use for interoperability between our DS-theoretic pragmatic reasoning component and our probabilistic reference resolution component.

We can use Dempster-Shafer Theory to represent the uncertainty of an event E using the interval $[Bl(E), Pl(E)]$. $Bl(E)$ and $Pl(E)$ are the *belief* and *plausibility* of E: lower and upper bounds on $P(E)$ such that $0 \leq Bl(E) \leq P(E) \leq Pl(E) \leq 1$. The *width* of this uncertainty interval $(Pl(E) - Bl(E))$ indicates the degree of *ignorance* one has regarding event E.

We thus take the following DS-theoretic approach. Let $\Theta = \{\theta_0, \dots, \theta_n\}$ be a *Frame of Discernment (FoD)* where each θ_i is a mutually exclusive singleton hypotheses described by ψ_i. Let $m(\cdot) : 2^\Theta \rightarrow [0, 1]$ be a *basic belief assignment* which assigns to each θ_i a mass:

$$\frac{1}{Z}P(\Phi_i \mid \Gamma_i, \Lambda_i), \quad (1)$$
where
$$Z = \sum_{j=0}^{|\Theta|} P(\Phi_j \mid \Gamma_j, \Lambda_j).$$

As mass is only assigned to singleton sets, $Bl(\theta_i) = Pl(\theta_i) = m(\theta_i)$. The confidence interval associated with each hypothesis according to this mass assignment is identical to $[Bl(\Gamma_i, \Lambda_i \mid \Phi_i), Pl(\Gamma_i, \Lambda_i \mid \Phi_i)]$ as calculated using Heendeni et al.'s (2016) DS-theoretic equivalent to Bayes' Rule (Eq. 2), assuming a uniform prior distribution $Bl(\Gamma, \Lambda) = Pl(\Gamma, \Lambda) = \frac{1}{|\Theta|}$.

$$\begin{aligned} Bl(A|B) &\geq \frac{Bl(B|A)Bl(A)}{Bl(B|A)Bl(A) + Pl(B|\bar{A})Pl(\bar{A})}; \\ Pl(A|B) &\leq \frac{Pl(B|A)Pl(A)}{Pl(B|A)Pl(A) + Bl(B|\bar{A})Bl(\bar{A})}. \end{aligned} \quad (2)$$

Before we move on, it is important to note that hypotheses with probabilities below a given threshold are pruned out during the resolution process, as described in our previous work (e.g., Williams et al. (2016)). This has the effect of concentrating extra probability mass in the remaining hypotheses, leading, respectively, to higher beliefs and plausibilities.

The result of the above calculations is a Frame of Discernment whose singleton hypotheses can be described by the logical conjunctions (i.e., BUSSes) $\psi_0 \ldots \psi_n$. However, the next component in the DIARC NL Pipeline (i.e., the pragmatic reasoning component) only uses the utterance form, and not the supplemental semantics, and there may be multiple hypotheses in Θ that have the same utterance form but different supplemental semantics.

As an example, if Bob had said "Grab the medkit *that is near the book*", and one candidate medkit (o_1) is actually near two books (o_2 and o_3), we could have two hypotheses which can be described by BUSSes that have the same utterance form (e.g. $Instruct(b, s, grab(s, o_1))$) but different supplemental semantics (e.g., $\{medkit(o_1) \wedge book(o_2) \wedge near(o_1, o_2)\}$ vs $\{medkit(o_1) \wedge book(o_3) \wedge near(o_1, o_3)\}$. We thus cluster these hypotheses into sets $C_0, \ldots, C_n$ such that all hypotheses associated with each set are described by BUSSes that have the same utterance form. As an example, if we have three singleton hypotheses $\{\theta_1, \theta_2\, \theta_3\}$, and ψ_1 and ψ_2 have the same utterance form, $C = \{\{\theta_1, \theta_2\}, \{\theta_3\}\}$.

We can now split our Frame of Discernment Θ into a set of $|C|$ "binary" FoDs, one for each cluster C_i. Each binary FoD itself has two hypotheses: (1) that the utterance form describing all hypotheses in cluster C_i *does* represent what was communicated, and (2) that it does not. This splitting has no theoretical ramifications, but facilitates easier integration with our pragmatic inference component. Because each cluster is mutually exclusive from all other clusters, each binary FoD can be represented entirely by the *bound utterance structure*:

$$\langle utterance(\psi_i), Bl(\{C_{i_0} \ldots C_{i_n}\}), Pl(\{C_{i_0} \ldots C_{i_n}\})\rangle.$$

Suppose $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and $\Psi = \{\psi_1, \psi_2, \psi_3\}$, where

$$\begin{aligned} \psi_1 = (&QuestionYN(b, s, can(s, grab(s, o_1))) \\ &\wedge medkit(o_1) \wedge book(o_2) \wedge near(o_1, o_2)), \\ \psi_2 = (&QuestionYN(b, s, can(s, grab(s, o_1))) \\ &\wedge medkit(o_1) \wedge book(o_3) \wedge near(o_1, o_3)), \\ \psi_3 = (&QuestionYN(b, s, can(s, grab(s, o_4))) \\ &\wedge medkit(o_4) \wedge book(o_2) \wedge near(o_4, o_2)), \end{aligned}$$

and assume the example basic belief assignment shown in the following table:

Hypothesis	Mass	Bl	Pl
$\emptyset$	0.0	0.0	0.0
$\{\theta_1\}$	0.2	0.2	0.2
$\{\theta_2\}$	0.3	0.3	0.3
$\{\theta_3\}$	0.5	0.5	0.5
$\{\theta_1, \theta_2\}$	0.0	0.5	0.5
$\{\theta_2, \theta_3\}$	0.0	0.8	0.8
$\{\theta_3, \theta_1\}$	0.0	0.7	0.7
$\{\theta_1, \theta_2, \theta_3\}$	1.0	1.0	1.0

Because ψ_1 and ψ_2 have the same utterance form, $C = \{\{\theta_1, \theta_2\}, \{\theta_3\}\}$. From this, the following set of bound utterance structures will be created:

$$\begin{aligned} \{&\langle QuestionYN(b, s, can(s, grab(s, o_1))), \\ &\qquad Bl(\{\theta_1, \theta_2\}), Pl(\{\theta_1, \theta_2\})\rangle, \\ &\langle QuestionYN(b, s, can(s, grab(s, o_4))), \\ &\qquad Bl(\{\theta_3\}), Pl(\{\theta_3\})\rangle\} = \\ \{&\langle QuestionYN(b, s, can(s, grab(s, o_1))), 0.5, 0.5\rangle \\ &\langle QuestionYN(b, s, can(s, grab(s, o_4))), 0.5, 0.5\rangle\} \end{aligned}$$

The set of bound utterance structures is sent to DIARC's DS-theoretic pragmatic reasoning component, which uses contextual knowledge to determine the intentions underlying these utterances (Williams et al. 2015). The pragmatic reasoning component produces a set of intentional structures $\langle I, Bl(I), Pl(I)\rangle$. If the difference between $Bl(I)$ and $Pl(I)$ is sufficiently large, or if $\frac{Pl(I)-Bl(I)}{2}$ is sufficiently close to 0.5, (assessed using Núñez et al.'s uncertainty measure (2013), shown in Eq. 3), intention I is deemed "uncertain" and in need of clarification.

$$1 + \frac{\beta}{K} log_2 \frac{\beta}{K} + \frac{1-\alpha}{K} log_2 \frac{1-\alpha}{K} \quad (3)$$
where $K = 1 + \beta - \alpha$.

If there are multiple intentions in need of clarification, the agent formulates an intention-to-know (*itk*) which intention is correct. This *itk* is denoted $itk(s, or(i_0, i_1, \ldots, i_n))$. We currently only handle situations with four or fewer possible interpretations. In future work, we plan to check for cases with five or more interpretations *before* they are sent through pragmatic reasoning; in such cases, a more general clarification request should be immediately generated.

Previously this *itk* only captured *intentional uncertainty* (e.g., when someone says "The commander needs a medical kit", it's possible they intend for the speaker to retrieve a medical kit for the commander, but it's also possible they intend for the speaker to inform them of

where to find a medical kit). Because the pragmatic inference process now receives a set of candidate utterance forms, each of which may have different argument bindings, this process thus acknowledges ambiguity, and thus captures *referential uncertainty* as well.

Before we move on, we would like to point out that that because DIARC's reference resolution component handles *open worlds*, instances in which interlocutors refer to previously unknown entities do not automatically generate clarification requests. For example, if the robot is told "Go to the room at the end of the hall" and did not previously know of a room at the end of the hall, it will not ask for clarification, but will rather hypothesize a new location, and carry on.

We do not regard such situations as referentially ambiguous. Here, the robot knows what entity is being referred to: a previously unknown room at the end of the hall. It may, of course, be valuable for the robot to ask for more information about this location, but we believe such a decision is not appropriate at the stage of processing we discuss in this paper.

Decision to Communicate

Currently, any such formulated *itk* is asserted into the robot's knowledge base, automatically triggering the decision to communicate this intention. Once it is acceptable for the robot to accept the conversational turn (as decided by a turn-taking algorithm), the robot will find this *itk* in its knowledge base and automatically decide to communicate it, passing the *itk* to the pragmatic *generation* component for processing.

Utterance Choice

During this stage, a robot must determine a contextually appropriate way to formulate its intention as a set of logical formulae. In DIARC, this is accomplished by the *pragmatic generation* component, which uses a set of DS-theoretic pragmatic rules. Each such rule maps an utterance to an intention under a particular context (these rules are also used for pragmatic understanding) (Williams et al. 2015). Using DS-theoretic logical operators, the pragmatic generation component is able to determine a set of candidate utterance forms, each of which is then forward-simulated through pragmatic inference in order to ensure that the agent does not accidentally communicate anything it does not actually believe to be true as a side effect of communicating its primary illocutionary point. The best candidate utterance is then sent to NLG for surface realization.

Because typical NLG systems do not need to account for social and dialogue context, this stage is not typically included. In contrast, NLG frameworks typically include a *document structuring* (c.f. (Reiter, Dale, and Feng 2000)) stage in which the agent determines the order in which to convey multiple utterances. Because situated clarification request generation typically only involves a single utterance, we do not currently handle this step. However, this will be an important topic for future work, since a robot may occasionally need to preface a clarification request by stating, for example, what aspects of an utterance it *did* understand.

Surface Realization

This stage subsumes facets of the lexical choice, referring expression generation, and realisation stages of Dale and Reiter's framework. While NLG capabilities have been previously integrated into the DIARC architecture, and even been used for clarification request generation (e.g., (Williams et al. 2015)), we have not yet implemented the Referring Expression Generation mechanisms necessary for robust surface realization. That is, while DIARC's NLG component can craft surface realizations for utterance forms such as $Statement(s, b, would\_like(s, medkit))$ (that is, a statement from an agent to bob that the agent would like a medkit), it does not yet handle utterance forms such as $Statement(s, b, need(s, obj_{34}))$, where obj_{34} may indeed be a medkit, but it is up to the agent to decide how best to describe it. We plan to integrate such mechanisms in future work.

Speech Synthesis

In DIARC, speech synthesis is performed using the open source MaryTTS (Schröder and Trouvain 2003) library.

Demonstration

To demonstrate the operation of the implemented framework stages, we present a proof-of-concept interaction that occurs in a simulated environment.

Architecture Configuration

For this interaction, we used one configuration of the DIARC Architecture. In addition to components responsible for the simulation of a Pioneer robot within an office environment, our configuration used the following components (see Fig. 2): ASR (which performs simulated speech recognition), NLP (which uses the C&C parser within a GH-theoretic framework), POWER (which performs reference resolution), AGENTS,SPEX and OBJECTS (POWER Consultants (c.f. (Williams and Scheutz 2015)) providing information about people, places, and things), DIALOGUE (which, performs dialogue management, and includes a pragmatic reasoning component as a submodule), BELIEF (which allows DIALOGUE to assess its current context), and ACTION (which performs goal and action management). The interaction begins with the speaker saying to the robot "I would like the medkit." ASR sends this to NLP, which parses the utterance into a dependency tree, from which it extracts root semantic content $would(X1, like(X1, X2))$, with utterance type $Statement$, additional semantic content $\{speaker(X1) \wedge medkit(X2)\}$, and presumed cognitive statuses $\{X1 \rightarrow definite, X2 \rightarrow definite\}$. Using this information, POWER searches for the referents to bind to $X1$ and $X2$; for $X1$, POWER finds a single probable candidate: $agents_1$, with probability 1.0; for $X2$, two candidates are found: $objects_1$, with probability of satisfaction 0.82, and $objects_2$, with probability of satisfaction 0.92. These bindings are then used to create

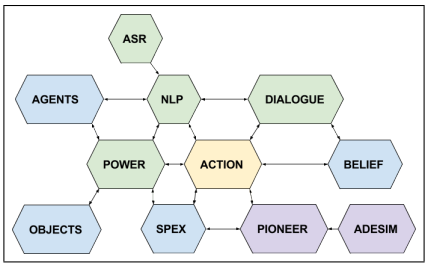

Figure 2: *Architecture Diagram.* Knowledge base components are depicted in blue; linguistic components are depicted in green; simulation components are depicted in purple; the action manager is depicted in yellow. The main contribution of this paper is the integration of the POWER and DIALOGUE components.

the following bound utterances[2]:

$$\{Statement(bob, self, would(bob, like(bob, objects_1))),\\ Statement(bob, self, would(bob, like(bob, objects_2)))\}$$

with corresponding probabilities[3] 0.82 and 0.92, respectively. These are normalized (see Eq. 1) and used to create DS-theoretic bound utterance structures, which are passed to DIALOGUE:

$$\{\langle Statement(bob, self, would(bob, like(bob, objects_1))),\\ 0.471, 0.471\rangle,\\ \langle Statement(bob, self, would(bob, like(bob, objects_2))),\\ 0.529, 0.529\rangle\}$$

The pragmatic reasoning component possess the rule:

$$\langle Statement(X, Y, would(Z, like(Z, W))) \Rightarrow \\ goal(Y, bring(Y, W, Z)), 0.9, 0.99\rangle, \quad (4)$$

indicating that the robot is between 90 and 99% confident in the rule; because the antecedent of this rule matches the utterance form of each bound utterance structure, uncertain Modus Ponens is applied in both cases, producing the set of intentional structures:

$$\{\langle goal(self, bring(self, objects_1, bob)), 0.424, 0.576\rangle,\\ \langle goal(self, bring(self, objects_2, bob)), 0.476, 0.524\rangle\}$$

Note that at this point, belief no longer equals plausibility: while the robot may not have encoded any ignorance with respect to what utterance was heard, ignorance encoded with respect to the context and rules the robot uses for pragmatic inference are reflected by ignorance now encoded with respect to the rules' consequents, thus painting a better picture of how much the robot truly knows about its interlocutor's intentions.

Nunez' uncertainty rule (see Eq. 3) determines that both of these intentions are highly uncertain. DIALOGUE thus determines its own intention to know which is correct, encoded as the structure:

$$\langle itk(self, or(goal(self, bring(self, objects_1, bob)),\\ goal(self, bring(self, objects_2, bob)))), 1.0, 1.0\rangle$$

[2]Here, $agent_1$ is changed to the name of that agent for the sake of dialogue processing.

[3]All beliefs and plausibilities in this section are rounded.

To decide how to communicate this intention, the bound utterance semantic structure is passed through the pragmatic reasoning component in reverse (Williams et al. 2015), using a rule of the form:

$$\langle QuestionWH(X, Y, or(Z, W)) \Rightarrow \\ itk(X, or(Z, W)), 0.95, 0.95\rangle, \quad (5)$$

Our approach allows recursive generation, allowing Eq. 5 to be chained with Eq. 4 to produce:

$$QuestionWH(self, bob, or(would(bob, like(bob, objects_1)),\\ would(bob, like(bob, objects_2)))).$$

At this point, we would ideally send this utterance form to our NLG component for generation of referring expressions for "bob", "$object_1$" and "$object_2$". As previously discussed, this will be a point for future work.

Conclusion

We have presented an HRI-oriented framework for clarification request generation, and shown how the first three stages of this framework as implemented in the DIARC architecture can identify and handle both pragmatic and referential ambiguity, both theoretically and in practice on a simulated robot. While this demonstration serves as proof-of-concept of the capabilities afforded by this integration effort, a full evaluation will clearly be necessary. Once all stages of the proposed framework have been implemented, we plan to run an extrinsic evaluation to determine how the extent to which the proposed algorithm benefits human-robot teaming in realistic HRI scenarios.

While the primary contributions of this paper is the finding that a pragmatic reasoning framework can track and address *referential* ambiguity, the work presented in this paper is also novel with respect to its integration of DS-theoretic and Bayesian theoretic architectural components. Because components of a robot architecture may often use different uncertainty frameworks, it is important for us to develop theoretically justified mechanisms for integrating such components. The presented approach provides one such technique for integration; in future work, we would like to examine others, as well as techniques to allow information to appropriately flow in the other direction (i.e., from DS-theoretic to Bayesian components).

There are several other extensions we would like to make in the near future. First, we must extend our approach in order to allow for more general questions to be asked. While prior research has shown that people prefer robots to enumerate options, there may be cases when it is necessary to ask a more general question, such as when there are a very large number of possible candidates for resolution. Future work must also involve integration of the REG capabilities required for the fourth stage of the proposed NLG framework. This will be the immediate focus of future work.

Acknowledgments

This work was in part funded by grant N00014-14-1-0149 from the US Office of Naval Research.

References

Brown, P. 1987. *Politeness: Some Universals in Language Usage*, volume 4. Cambridge University Press.

Clark, H. H. 1996. *Using language*. Cambridge university press.

Deits, R.; Tellex, S.; Thaker, P.; Simeonov, D.; Kollar, T.; and Roy, N. 2013. Clarifying commands with information-theoretic human-robot dialog. *Journal of Human-Robot Interaction* 2(2):58–79.

Fong, T.; Thorpe, C.; and Baur, C. 2001. Collaboration, dialogue and human-robot interaction, 10th international sumposium of robotics research (lorne, victoria, australia). In *Proceedings of the 10th International Symposium of Robotics Research*.

Garoufi, K., and Koller, A. 2014. Generation of effective referring expressions in situated context. *Language, Cognition and Neuroscience* 29(8):986–1001.

Gundel, J. K.; Hedberg, N.; and Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *language* 274–307.

Heendeni, J. N.; Premaratne, K.; Murthi, M.; Uscinski, J.; and Scheutz, M. 2016. A generalization of Bayesian inference in the Dempster-Shafer belief theoretic framework. In *Proceedings of the 2016 Intenational Conference on Information Fusion*.

Hemachandra, S.; Walter, M. R.; and Teller, S. 2014. Information theoretic question asking to improve spatial semantic representations. In *2014 AAAI Fall Symposium Series*.

Kruijff, G.-J. M.; Zender, H.; Jensfelt, P.; and Christensen, H. I. 2006. Clarification dialogues in human-augmented mapping. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 282–289. ACM.

Kruijff, G.-J. M.; Brenner, M.; and Hawes, N. 2008. Continual planning for cross-modal situated clarification in human-robot interaction. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 592–597. IEEE.

Marge, M., and Rudnicky, A. I. 2015. Miscommunication recovery in physically situated dialogue. In *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 22.

Núñez, R. C.; Dabarera, R.; Scheutz, M.; Briggs, G.; Bueno, O.; Premaratne, K.; and Murthi, M. N. 2013. DS-Based Uncertain Implication Rules for Inference and Fusion Applications. In *16th International Conference on Information Fusion*.

Purver, M.; Ginzburg, J.; and Healey, P. 2003. On the means for clarification in dialogue. In *Current and new directions in discourse and dialogue*. Springer. 235–255.

Purver, M. 2004. Clarie: The clarification engine. In *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue (Catalog)*, 77–84. Citeseer.

Reiter, E.; Dale, R.; and Feng, Z. 2000. *Building natural language generation systems*, volume 33. MIT Press.

Rosenthal, S.; Veloso, M.; and Dey, A. K. 2012. Is someone in this office available to help me? *Journal of Intelligent & Robotic Systems* 66(1-2):205–221.

Scheutz, M.; Schermerhorn, P.; Kramer, J.; and Anderson, D. 2007. First steps toward natural human-like HRI. *Autonomous Robots* 22(4):411–423.

Schröder, M., and Trouvain, J. 2003. The german text-to-speech synthesis system mary: A tool for research, development and teaching. *International Journal of Speech Technology* 6(4):365–377.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton University Press.

Stirling, A. 2010. Keep it complex. *Nature* 468(7327):1029–1031.

Stoyanchev, S.; Liu, A.; and Hirschberg, J. 2013. Modelling human clarification strategies. In *Proceedings of SIGDIAL*, 137–141.

Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M. R.; Banerjee, A. G.; Teller, S.; and Roy, N. 2011. Approaching the symbol grounding problem with probabilistic graphical models. *AI magazine* 32(4):64–76.

Tellex, S.; Thaker, P.; Deits, R.; Simeonov, D.; Kollar, T.; and Roy, N. 2013. Toward information theoretic human-robot dialog. *Robotics* 409.

Tenbrink, T.; Ross, R. J.; Thomas, K. E.; Dethlefs, N.; and Andonova, E. 2010. Route instructions in map-based human–human and human–computer dialogue: A comparative analysis. *Journal of Visual Languages & Computing* 21(5):292–309.

Traum, D. R. 1994. A computational theory of grounding in natural language conversation. Technical report, DTIC Document.

Williams, T., and Scheutz, M. 2015. POWER: A domain-independent algorithm for probabilistic, open-world entity resolution. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Williams, T., and Scheutz, M. 2016. A framework for resolving open-world referential expressions in distributed heterogeneous knowledge bases. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*.

Williams, T.; Nunez, R. C.; Briggs, G.; Scheutz, M.; Premaratne, K.; and Murthi, M. N. 2014. A dempster-shafer theoretic approach to understanding indirect speech acts. In *Advances in Artificial Intelligence*.

Williams, T.; Briggs, G.; Oosterveld, B.; and Scheutz, M. 2015. Going beyond literal command-based instructions: Extending robotic natural language interaction capabilities. In *Proceedings of Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Williams, T.; Acharya, S.; Schreitter, S.; and Scheutz, M. 2016. Situated open world reference resolution for human-robot dialogue. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction*.