

Facilitating Mental Modeling in Collaborative Human-Robot Interaction through Adverbial Cues

Gordon Briggs and Matthias Scheutz
Human-Robot Interaction Laboratory
Tufts University, Medford, MA 02155, USA
{gbriggs, mscheutz}@cs.tufts.edu

Abstract

Mental modeling is crucial for natural human-robot interactions (HRI). Yet, effective mechanisms that enable reasoning about and communication of mental states are not available. We propose to utilize adverbial cues, routinely employed by humans, for this goal and present a novel algorithm that integrates adverbial modifiers with belief revision and expression, phrasing utterances based on Gricean conversational maxims. The algorithm is demonstrated in a simple HRI scenario.

1 Introduction

Advances in robotics and autonomous systems are paving the way for the development of robots that can take on increasingly complex tasks without the need of minute human supervision. As a result of this greater autonomy, the interaction styles between humans and robots are slowly shifting from those of humans micromanaging robot behaviors (e.g., via remote controls) to more higher-level interactions (e.g., verbal commands) which are required for many mixed initiative tasks where humans and robots work together in teams (e.g., in search and rescue missions). In order for these joint human-robot interactions to be productive and efficient, robots must have the ability to communicate in *natural* and *human-like* ways (Scheutz et al., 2007). Natural human-like communication in robots, however, requires us to tackle several challenges, including the development of robust natural language (NL) competencies and the ability to understand and utilize a variety of affective, gestural, and other non-

linguistic cues that are indicative of the interlocutor's mental states. Hence, natural human-like interaction also requires the construction and maintenance of mental models of other agents, especially in the context of collaborative team tasks where actions among multiple agents must be coordinated, often through natural language dialogues.

Several recent efforts are aimed at endowing robots with natural language processing capabilities to allow for verbal instructions as a first step (e.g., (Brenner, 2007; Dzifcak et al., 2009; Kress-Gazit et al., 2008; Rybski et al., 2007; Kollar et al., 2010)). Independently, user modeling has been extensively explored in order to generate more natural and productive human-machine interactions (Kobsa, 2001), including adapting the natural language output of dialogue systems based on mental models of human-users (Wahlster and Kobsa, 1989). However, there is currently no integrated robotic architecture that includes explicit mechanisms for efficiently conveying natural language information about the robot's "mental states" (i.e., beliefs, goals, intentions) to a human teammate. Yet, such mechanisms are not only desirable to make the robot's behavior more intuitive and predictable to humans, but can also be crucial for team success (e.g., quick updates on goal achievement or early corrections of wrong human assumptions).

We propose a novel integrated belief revision and expression algorithm that allows robots to track and update the beliefs of their interlocutors in a way that respects Gricean maxims about language usage. The algorithm explicitly models and updates task-relevant beliefs and intentions of all participating

agents. Whenever a discrepancy is detected between a human belief (as implied in a natural language expression uttered by the human) and the robot’s mental model of the human, the robot generates a natural language response that corrects the discrepancy in the most effective way. To achieve effectiveness, the robot uses linguistic rules about the pragmatic implications of adverbial modifiers like “yet”, “still”, “already”, and others that are used by humans to effectively communicate their beliefs and intentions.

The rest of the paper is organized as follows. We start with a motivation of our approach based on Gricean maxims. Then, we introduce formalizations of linguistic devices that humans use to generate effective task-based dialogue interactions and present our algorithm for generating appropriate utterances in response to human queries. Next we use a simple remote human-robot interaction scenario to demonstrate the operation of the algorithm, followed by a discussion and summary of our contributions.

2 Motivation

Joint activity often requires agents to monitor and keep track of each others’ mental states to ensure effective team performance. For example, searchers during rescue operations in disaster zones typically coordinate their (distributed and remote) activities through spoken natural language interactions via wireless audio links to keep team members informed of discoveries and plans of other team members. Coordination as part of joint activities requires two important processes in an agent: (1) building and maintaining a *mental model* of the other agents’ beliefs and intentions (based on perceived, communicated, and inferred information), which is critical for situational awareness (Lison et al., 2010); and (2) actively supporting the maintenance of others’ mental models of oneself (e.g., by proactively communicating new information to the other agents in ways that will allow them to update their mental models).

Cohen et al. (1990), for example, discuss the necessity of various communicative acts that serve to synchronize agent belief-models. These communicative acts include both linguistic and non-linguistic cues, such as utterances of confirmation (“okay.”) or signals that indicate intention (putting on a turn-signal). In addition to utilizing explicit

cues to synchronize belief-models, humans employ various other mechanisms to convey information about one’s own belief-state, in particular, various linguistic devices. A simple, but very powerful linguistic mechanism is the use of *adverbial cues*.

Consider a scenario where one agent wants to know the location of another agent, e.g., whether the agent is at home. A straightforward way to obtain this information is to simply ask “Are you at home?” The other agent can then answer “yes” or “no” accordingly. Now, suppose the first agent knew that the second agent was planning to be at home at some point. In that case, the agent might ask “Are you at home yet?” Note that semantically both questions have the same meaning, but their pragmatic implications are different as the second implies that that agent 1 knows that agent 2 was planning to be at home, while no such implication can be inferred from the first query. Conversely, suppose that agent 2 responded “not yet” in the first example (instead of “no”). While the semantic meaning is the same as “no”, “not yet” communicates to agent 1 that agent 2 has the goal to be home. In general, adverbs like “yet” can be used to convey information about one’s (or somebody else’s) beliefs concerning mutually-recognized goals and intentions. Not surprisingly, humans use them regularly and with ease to aid their interlocutors with maintaining an accurate model of their beliefs and goals.

The challenges that need to be addressed to allow robots to have the above kinds of linguistic exchanges are: (1) how to formalize the functional roles of adverbial modifiers in different sentence types, and how to use the formalized principles to (2) perform belief updates and (3) generate effective natural language responses that are natural, succinct, and complete. To tackle these three challenges, we turn to Gricean principles that have long been used in pragmatics as guiding principles of human communicative exchanges.

3 NL Understanding and Generation

Grice (1975) proposed four general principles to aid in the pragmatic analysis of utterances. Phrased as rules, it is unsurprising that they have been used as an inspiration for NL generation systems before. Dale and Reiter (1995) have enlisted the maxims in

their design of an algorithm to generate referring expressions, while others have cited Gricean influence in utterance selection for intelligent tutor systems (Eugenio et al., 2008). The particular maxims we considered are the maxims of *quality* (G1), *quantity* (G2), and *relevance* (G3): (G1) requires one to not say what one believes is false or for which one lacks adequate evidence; (G2) requires one to make contributions as informative as necessary for the current purposes of the exchange, but not more informative; and (G3) tersely states “be relevant.”

Our approach to belief-model synchronization and utterance selection is based on the above maxims and attempts to select the most appropriate response to another agent’s query based on relevance of semantic content. It uses speech pragmatic meaning postulates for linguistic devices such as adverbial modifiers to search for a succinct and natural linguistic representation that captures the intended updates. Rather than explicitly communicating each and every proposition that needs to be communicated to a human to allow the person to update their mental model of the robot, the algorithm makes heavy use of “implied meanings”, i.e., propositions that humans will infer from the way the information is phrased linguistically. This allows for much shorter messages to be communicated than otherwise possible and addresses the second maxim of quantity.

3.1 Formalizing pragmatic implications

We start by introducing four types of sentences as they are found in typical dialogue interactions: *statements* (expressed through declarative sentences), *questions* (expressed through interrogative sentences), *commands* (expressed through imperative sentences) and *acknowledgments* (expressed through words like “okay”, “yes”, “no”, etc.). For simplicity, we restrict the discussion to one predicate $at(\alpha, \lambda)$ which states that agent α is in location λ .

3.1.1 Statements

We will use the form $Stmt(\alpha, \beta, \phi, \mu)$ to express that agent α communicates ϕ to agent β using adverbial modifiers in a set μ . For example, $Stmt(A_2, A_1, \neg at(A_2, home), yet)$ means that agent A_2 tells A_1 that it is not at home yet. Note

that we are indifferent about the exact linguistic representation of ϕ here as the goal is to capture the pragmatic implications.

If α informs β that it is at λ without any adverbial modifiers or additional contextual information, then we can assume using (G1) that α is indeed at that location:

$$[[Stmt(\alpha, \beta, at(\alpha, \lambda), \{\})]]_c := at(\alpha, \lambda) \quad (1)$$

Here we use $[[\cdot]]_c$ to denote the “pragmatic meaning” of an expression in context c , which includes task, goal, belief and discourse aspects. Next, we inductively define the pragmatic meanings for several adverbial modifiers “still”, “already”, “now”, and “not yet” (the meanings of compound expressions such as $at(\alpha, \lambda_1) \wedge \neg at(\alpha, \lambda_2)$ are defined recursively in the usual way).

If α states that it is “still” at λ , one can infer that α is at λ and that α will not be at λ at some point in the future:

$$[[Stmt(\alpha, \beta, at(\alpha, \lambda), \{still\})]]_c := [[Stmt(\alpha, \beta, at(\alpha, \lambda), \{\})]]_c \wedge Future(\neg at(\alpha, \lambda)) \quad (2)$$

If α states that it is “already” at λ , one can infer that α is at λ and that α had a goal (expressed via the “ G ” operator) to be at λ at some point in the past:

$$[[Stmt(\alpha, at(\alpha, \lambda), \{already\})]]_c := [[Stmt(\alpha, \beta, at(\alpha, \lambda), \{\})]]_c \wedge Past(G(\alpha, at(\alpha, \lambda))) \quad (3)$$

If α states that it is “now” at λ , one can infer that α is at λ and that α had not been at λ at some point in the past:

$$[[Stmt(\alpha, \beta, at(\alpha, \lambda), \{now\})]]_c := [[Stmt(\alpha, \beta, at(\alpha, \lambda), \{\})]]_c \wedge Past(\neg at(\alpha, \lambda)) \quad (4)$$

If α states that it is “not...yet” at λ , one can infer that α is not at λ , but has an intention to be at λ .

$$[[Stmt(\alpha, \beta, \neg at(\alpha, \lambda), \{yet\})]]_c := \neg at(\alpha, \lambda) \wedge G(\alpha, at(\alpha, \lambda)) \quad (5)$$

Even in our limited domain, one must be cognizant of the ambiguities that arise from how adverbial cues are deployed. In addition to the simple presence of an adverbial cue, the location of the adverb in a sentence and prosodic factors may affect the intended meaning of the utterance. For instance, consider the statements: (a) I am now at λ ; (b) I am

at λ now; (c) I am still at λ ; and (d) I am *still* at λ . Statement (a) is a simple situational update utterance as described above, while (b) could be construed as a statement akin to “I am already at λ ”. Statement (d) could be interpreted as additionally signaling the frustration of the agent, beyond conveying the information from (c).

It should also be noted that our analysis of these adverbial cues is to be understood in the limited context of these simple task-related predicates (e.g. $at(\alpha, \lambda)$). Formal definition of these adverbial cues in general cases is beyond the scope of this paper. For instance, “yet” could be used in a context when the predicate is not intended by the agent to which it applies (e.g. “Has Bill been fired yet?”). In this case, it would probably be incorrect to infer that the agent Bill had a goal to be fired. Instead an inference could be made regarding the probabilistic judgments of the interlocutors regarding the topic agent’s future state. However, in the context of this paper, it is assumed that “yet” is used in the context of goals intended by agents.

3.1.2 Questions

Here we will limit the discussion to two question types, the “where” question (regarding locations) and simple “Yes-No” questions.

If α asks β about its location in the general sense (“where are you?”), then one can infer that α has an intention to know (expressed via the “IK” operator, see (Perrault and Allen, 1980)) where β is located:

$$[[Ask_{loc}(\alpha, \beta, \{\})]]_c := IK(\alpha, at(\beta, \lambda)) \quad (6)$$

for some λ .

If α asks β whether it is at λ , then one can infer that α has an intention to know whether β is at λ :

$$[[Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \{\})]]_c := IK(\alpha, at(\beta, \lambda)) \quad (7)$$

If β is asked by α whether it is “still” at λ , β can infer that α believes (expressed via the “B” operator) that β is currently at λ :

$$\begin{aligned} [[Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \{still\})]]_c &:= & (8) \\ [[Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \{\})]]_c \wedge B(\alpha, at(\beta, \lambda)) \end{aligned}$$

If β is asked by α whether it is at λ “yet”, β can infer that α believes that β has a goal to be at λ :

$$\begin{aligned} [[Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \{yet\})]]_c &:= & (9) \\ [[Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \{\})]]_c \wedge B(\alpha, G(\beta, at(\beta, \lambda))) \end{aligned}$$

3.1.3 Question-Answer Pairs

Next, we consider how discourse context as provided by question-answer pairs can further specify the pragmatic implications.

If α asks β whether it is at λ with any set of adverbial modifiers μ (i.e., $Prior(Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \mu)) \in c$), and β responds by stating that it is “still” at λ , then one can infer that α has the belief that β was at λ in the recent past:

$$\begin{aligned} [[Stmt(\beta, \alpha, at(\beta, \lambda), \{still\})]]_c &:= & (10) \\ [[Stmt(\beta, \alpha, at(\beta, \lambda), \{\})]]_c \\ \wedge B(\alpha, RecPast(at(\beta, \lambda))) \end{aligned}$$

where $Prior(Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \mu)) \in c$. Also, $RecPast(\phi)$ denotes that ϕ was true in the recent past, as distinct from ϕ holding at some arbitrary point in the past (i.e. $Past(\phi)$). This distinction is necessary as it only makes sense to use the adverbial cue at this point if agent α believed $at(\beta, \lambda)$ at some relative and recent point in the past. Formalizing this would require keeping track of the points in time at which certain propositions are believed. To avoid committing to a particular temporal modeling system, we make the simplifying assumption that the $RecPast$ operator is not applied in rules (10) and (11), which is sufficient for the very simple interactions examined in this paper.

If α asks β whether it is at λ with any set of adverbial modifiers μ , and β responds by stating that it is “now” at λ , then one can infer that α has the belief that β is was not at λ in the recent past:

$$\begin{aligned} [[Stmt(\beta, \alpha, at(\beta, \lambda), \{now\})]]_c &:= & (11) \\ [[Stmt(\beta, \alpha, at(\beta, \lambda), \{\})]]_c \\ \wedge B(\alpha, RecPast(\neg at(\beta, \lambda))) \end{aligned}$$

where $Prior(Ask_{yn}(\alpha, \beta, at(\beta, \lambda), \mu)) \in c$.

3.1.4 Commands

We also briefly describe how command processing (which we have studied elsewhere in much greater detail (Dzifcak et al., 2009)) can be augmented with the inclusion of pragmatic meanings. If α orders β to travel to λ , then one can infer that α has a goal for β to be at λ and that α intends to know whether β has received its new goal:

$$\begin{aligned} [[Cmd(\alpha, \beta, at(\beta, \lambda), \{\})]]_c &:= & (12) \\ G(\alpha, at(\beta, \lambda)) \\ \wedge IK(\alpha, G(\beta, at(\beta, \lambda))) \end{aligned}$$

It would be an oversimplification to assume that the proposition $G(\beta, at(\beta, \lambda))$ is immediately understood by all listening agents. In order to generate the appropriate goal belief in the target agent, additional inference rules need to be considered. The following rule states that β will instantiate the goal $G(\beta, at(\beta, \lambda))$ when it believes α has the same goal and it believes $authority(\alpha, \beta)$, which denotes that α has command authority over β :

$$G(\alpha, at(\beta, \lambda)) \wedge authority(\alpha) \Rightarrow G(\beta, at(\beta, \lambda))$$

Other agents would have to wait for an acknowledgment that this inference has indeed taken place (as β could have not heard the initial command utterance). These acknowledgment utterances are described in the subsequent section.

3.1.5 Acknowledgments

Finally, we consider typical forms of acknowledgment. If α utters an acknowledgment (e.g., “OK.”) when the previous utterance was a positive statement of location by β , then one can infer α no longer has the intention to know β ’s location:

$$[[Ack(\alpha, \beta, \{\})]]_c := \neg IK(\alpha, at(\beta, \lambda)) \quad (13)$$

for some λ where for any M $Prior(Stmt(\beta, \alpha, at(\beta, \lambda), \{M\})) \in c$.

If α utters an acknowledgment (e.g., “OK.”) when the previous utterance was a command by β to be at λ , then one can infer that

$$[[Ack(\alpha, \beta, \{\})]]_c := G(\alpha, at(\alpha, \lambda)) \wedge G(\beta, at(\alpha, \lambda)) \wedge \neg IK(\beta, G(\alpha, at(\alpha, \lambda))) \quad (14)$$

where $Prior(Cmd(\beta, \alpha, at(\alpha, \lambda), \{M\})) \in c$ for any M .

We should note here that the distinction between explicitly not intending-to-know and the lack of an intention-to-know has been blurred in the above rules for the sake of simplicity. As described in the subsequent section, agent beliefs are removed when contradicted in the current system (i.e. $Remove(\phi, B_\alpha) \Leftrightarrow (\neg\phi) \in B_\alpha$). A more comprehensive belief update system should allow for a mechanism to remove beliefs without the need for explicit contradiction.

3.2 Agent Modeling and Belief Updates

Belief updates occur whenever an agent α receives an utterance Utt from another agent β in context c . First, $[[Utt]]_c$ is computed using the pragmatic principles and definitions developed in Section 3.1. For simplicity, we assume that agents adhere to the Gricean maxim of *quality* and, therefore, do not communicate information they do not believe. Hence, all propositions $\phi \in [[Utt]]_c$ are assumed to be true and to the extent that they are inconsistent with existing beliefs of α as determined by α ’s inference algorithm \Rightarrow_α^b , the conflicting beliefs are removed from the agent’s sets of beliefs Bel_{self} (b here denotes some finite bound on the inference algorithm, e.g., resources, computation time, etc.).¹

To model other agents hearing the utterance, agent α derives the set $B_\alpha B_\gamma = \{\psi | B(\gamma, \psi) \in Bel_{self}\}$ for all other agents $\gamma \neq \alpha$. The agent updates these belief sets by applying the same rules as it does to Bel_{self} .

It should be noted that these belief update rules are indeed simplifications designed to avoid the issue of resolving conflicting information from different sources. These belief update rules would be problematic, for instance, when agents have incorrect beliefs (and proceed to communicate them), as no method for belief disputation exists. For the purpose of illustrating the implementation and utility of adverbial cues, however, they should suffice. We set up our environment and rule sets such that the autonomous agent has perfect information about itself (specifically location), and no utterances exists to communicate propositions that are not about one-self.

3.3 Sentence Generation

Depending on the sentence type α received (and the extent to which meanings can be resolved, an issue we will not address in this paper), different response sentence types are appropriate (e.g., a yes-no ques-

¹Note that we are not making any assumption about a particular inference algorithm or its (as it will, in general, depend on the expressive power of the employed logic to represent meanings), only that if a contradiction can be reached using the inference algorithm, the existing belief needs to be removed (otherwise existing beliefs are taken to be consistent with the implications of the utterance). In our implemented system, we use a simplified version of the resolution inference principle.

tion requires a statement answering the question). The generation of an appropriate response proceeds in two steps. First, based on the agent’s current set of beliefs Bel_{self} , we determine the set of propositions Φ_{comm} that the agent has an interest in conveying. Second, we attempt to find the smallest utterance Utt given a set of pragmatic principles (as specified in Section 3.1) that communicates one or more of these propositions and implies the rest for recipient β .

3.3.1 What to say

In obtaining a set Φ_{comm} of propositions to communicate, α may obey the Gricean maxim of *quality* by adding a proposition ϕ to Φ_{comm} only if $\phi \in Bel_{self}$. The maxims of *relevance* and *quantity* are heeded by restricting believed propositions to be conveyed solely to those that either correct a false belief of β or provide β some piece of information it wants to know. Specifically, we find the set of all propositions used to correct false beliefs Φ_{rev} , defined as:

$$\psi \in \Phi_{rev} \Leftrightarrow \exists \beta, \phi : \\ B(\beta, \phi) \wedge \phi \in Bel_{self} \wedge (\psi \Rightarrow_{\alpha}^b \neg \phi)$$

The set of all propositions other agents want to know, Φ_{IK} , can be defined as:

$$\psi \in \Phi_{IK} \Leftrightarrow \exists \beta, \phi : \psi \in Bel_{self} \wedge \\ IK(\beta, \phi \in Bel_{self}) \wedge (\psi \Rightarrow_{\alpha}^b \phi \vee \psi \Rightarrow_{\alpha}^b \neg \phi)$$

The final set of propositions to convey is obtained by merging these two sets, $\Phi_{comm} = \Phi_{rev} \cup \Phi_{IK}$. Note that this set is always consistent because propositions are added to Φ_{rev} and Φ_{IK} if and only if they exist in Bel_{self} , which is maintained to be consistent.

3.3.2 How to say it

Once Φ_{comm} has been obtained, α must select potential utterances to produce. It starts by generating an initial set Utt_0 of utterances that in the present context c imply some subset of Φ_{comm} :

$$(u \in Utt_0) \Leftrightarrow \exists \Phi \in \Phi_{comm} \forall \phi \in \Phi : ([[u]]_c \Rightarrow_{\alpha}^b \phi)$$

Currently, this is achieved by searching through the set of all utterances defined by rules such as those

found in Section 3.1. Note that while this approach is feasible for our quite limited domain, more efficient methods for identifying candidate utterances must be developed as the number of understood utterances grows.

Applying the maxim of *quality*, this set can be pruned of all utterances that are defined by additional propositions that we either have no evidence for (“unsupported”) or explicitly believe to be false:

$$False(\phi) \Leftrightarrow \exists \psi : \psi \in Bel_{self} \wedge (\psi \Rightarrow_{\alpha}^b \neg \phi)$$

$$NoSupp(\phi) \Leftrightarrow \neg \exists \psi : \psi \in Bel_{self} \wedge (\psi \Rightarrow_{\alpha}^b \phi)$$

Using these conditions, we can generate a new subset of utterance candidates Utt_1 :

$$(u \in Utt_1) \Leftrightarrow \neg \exists \phi : ([[u]]_c \Rightarrow_{\alpha}^b \phi) \\ \wedge (False(\phi) \vee NoSupp(\phi))$$

Applying the maxim of *quantity*, utterances that revise or add the most beliefs to other agent belief-spaces ought to be favored:

$$RevBel(\beta, \phi) \Leftrightarrow \\ \exists \psi : B(\beta, \psi) \in B_{self} \wedge (\psi \Rightarrow_{\alpha}^b \neg \phi)$$

$$AddBel(\beta, \phi) \Leftrightarrow B(\beta, \phi) \notin Bel_{self}$$

Using these definitions, we can derive the “correction-score” of an utterance by counting the number of propositions $\phi \in [[u]]_c$ that revise or add a belief for β .

If multiple candidate utterances still exist at this point, we can again apply the maxim of *quantity* to favor utterances that convey the most (true) information. Because all definitions with false propositions have been eliminated, we can simply count the number of true propositions derived from the utterance, thereby favoring semantically richer utterances. At this point, if multiple candidate utterances are still available, the difference is of stylistic nature only and we may choose an arbitrary one. Note that the correct usage of adverbial modifiers emerges naturally from these rules as utterances that include inappropriate adverbs are removed in Utt_1 , while utterances that include appropriate adverbial cues are subsequently favored.

4 Case Study

We now demonstrate the operation of the proposed algorithm in a simple joint activity scenario where a robot (R) is located at nav-point 1 and correctly knows its location, having the initial belief-space $B_R = \{at(R, N1)\}$. The remote human operator starts by asking:

O: R, where are you?

R updates its beliefs based on this question:

$$\begin{aligned} u &:= \text{parse}(\text{"O: R, where are you?"}) \\ \rightarrow u &:= \text{Ask}_{loc}(O, R, \{\}) \\ [[u]]_c &:= \{IK(O, at(R, N1)), IK(O, at(R, N2)), \\ &IK(O, at(R, N3))\} \\ P_{contra} &:= \text{contradictedTerms}([[u]]_c, B_{self}) \\ B_R &:= (B_R - P_{contra}) + [[u]]_c \\ B_{RBO} &:= (B_{RBO} - P_{contra}) + [[u]]_c \end{aligned}$$

which yields a new belief-space:

$$\begin{aligned} B_R &:= \{at(R, N1), IK(O, at(R, N1)), \\ &IK(O, at(R, N2)), IK(O, at(R, N3)), \\ &B(O, IK(O, at(R, N1))), B(O, IK(O, at(R, N2))), \\ &B(O, IK(O, at(R, N3)))\} \end{aligned}$$

Next, R proceeds to respond. For compactness, we refer below to utterance candidates according to the index of the applicable rules from Section 3.1, so that u_{13} denotes $Ack(\alpha, \beta, \{\})$.

$$\begin{aligned} B_{RBO} &:= \{IK(O, at(R, N1)), IK(O, at(R, N2)) \\ &IK(O, at(R, N3))\} \\ \Phi_{rev} &:= \{\}; \Phi_{IK} := \{at(R, N1)\} \\ \Phi_{comm} &:= \{at(R, N1)\}; \\ \rightarrow Utt_0 &:= \{u_1, u_2, u_3, u_4\} \end{aligned}$$

R now has an initial set of candidate utterances, which it prunes using the rules from Section 3.3.2.

$$\begin{aligned} [[u_1]]_c &:= at(R, N1) \\ [[u_2]]_c &:= at(R, N1) \wedge \text{Future}(\neg at(R, N1)) \\ [[u_3]]_c &:= at(R, N1) \wedge \text{Past}(G(R, N1)) \\ [[u_4]]_c &:= at(R, N1) \wedge \text{Past}(\neg at(R, N1)) \\ \rightarrow Utt_1 &:= \{u_1\} \end{aligned}$$

Thus, R chooses the utterance of the form, $Stmnt(R, O, at(R, N1), \{\})$, and responds:

R: I am at N1.

Finally, R processes its own utterance so that it can update its beliefs according to rule (1):

$$\begin{aligned} B_R &:= \{at(R, N1), IK(O, at(R, N1)), \\ &IK(O, at(R, N2)), IK(O, at(R, N3)), \\ &B(O, IK(O, at(R, N1))), B(O, IK(O, at(R, N2))), \\ &B(O, IK(O, at(R, N3))), B(O, at(R, N1))\} \end{aligned}$$

When the operator responds:

O: Okay.

R also processes this acknowledgment to update its beliefs according to rule (13):

$$B_R := \{at(R, N1), B(O, at(R, N1))\}$$

R proceeds to respond, but finds that it has nothing to convey.

$$\begin{aligned} B_{RBO} &:= \{at(R, N1)\} \\ \Phi_{rev} &:= \{\}; \Phi_{IK} := \{\} \\ \Phi_{comm} &:= \{\}; \\ \rightarrow Utt_0 &:= \{\} \end{aligned}$$

Thus, R generates no utterance. Now let us suppose that R moves to N2, and enough time elapses such that the operator forfeits his/her conversational turn.

R then proceeds to generate an utterance.

$$\begin{aligned} B_R &:= \{at(R, N2), \text{Past}(at(R, N1))\} \\ B_{RBO} &:= \{at(R, N1)\} \\ \Phi_{rev} &:= \{at(R, N2)\}; \Phi_{IK} := \{\} \\ \Phi_{comm} &:= \{at(R, N2)\} \\ \rightarrow Utt_0 &:= \{u_1, u_2, u_3, u_4\} \\ [[u_1]]_c &:= at(R, N2) \\ [[u_2]]_c &:= at(R, N2) \wedge \text{Future}(\neg at(R, N2)) \\ [[u_3]]_c &:= at(R, N2) \wedge G(R, N2) \\ [[u_4]]_c &:= at(R, N2) \wedge \text{Past}(\neg at(R, N2)) \\ \rightarrow Utt_1 &:= \{u_1, u_4\} \end{aligned}$$

So, R must now resolve which of these candidate utterances to select by choosing the one that revises the most beliefs of O, or failing that, the one that has the most true propositions.

$$\begin{aligned} at(R, N2) &\Rightarrow \neg at(R, N1) \\ \rightarrow \text{NumRev}([[u_1]]_c) &:= 1; \text{NumRev}([[u_4]]_c) := 1; \\ \text{NumTrue}([[u_1]]_c) &:= 1; \text{NumTrue}([[u_4]]_c) := 2; \\ \rightarrow Utt_{final} &:= u_4 \end{aligned}$$

Thus, R chooses the utterance of the form, $Stmnt(R, O, at(R, N2), \{now\})$, and responds:

R: I am now at N2.

R again processes its own utterance to update its beliefs according to rule (4). If O then asks:

O: R, are you still at N2?

R updates its beliefs according to rule (10):

$$\begin{aligned} B_R &:= \{at(R, N2), B(O, at(R, N2)), \\ &\text{Past}(at(R, N1)), B(O, \text{Past}(at(R, N1))), \\ &IK(O, at(R, N2)), B(O, IK(O, at(R, N2)))\} \end{aligned}$$

Next, R generates a response:

$$\begin{aligned} \Phi_{rev} &:= \{\}; \Phi_{IK} := \{at(R, N2)\} \\ \Phi_{comm} &:= \{at(R, N2)\}; \\ \rightarrow Utt_0 &:= \{u_1, u_2, u_{10}, u_{11}\} \\ [[u_1]]_c &:= at(R, N2) \\ [[u_2]]_c &:= at(R, N2) \wedge \text{Future}(\neg at(R, N2)) \\ [[u_{10}]]_c &:= at(R, N2) \wedge B(O, at(R, N2)) \\ [[u_{11}]]_c &:= at(R, N2) \wedge B(O, \neg at(R, N2)) \\ \rightarrow Utt_1 &:= \{u_1, u_{10}\} \\ \rightarrow \text{NumRev}([[u_1]]_c) &:= 0; \text{NumRev}([[u_{10}]]_c) := 0; \\ \text{NumTrue}([[u_1]]_c) &:= 1; \text{NumTrue}([[u_{10}]]_c) := 2; \\ \rightarrow Utt_{final} &:= u_{10} \end{aligned}$$

replying with the utterance:

R: I am still at N2.

and processes its own utterance to update its beliefs according to rule (10). O's acknowledgment:

O: Okay.

causes R to update its beliefs according to rule (13):

$$B_R := \{at(R, N2), B(O, at(R, N2)), \\ Past(at(R, N1)), B(O, Past(at(R, N1)))\}$$

R does not generate a response as there are no beliefs to revise or intentions to know. Now suppose R moves back to N1, without O's knowledge, after which O commands:

O: R, go to N1.

R, updates its belief according to rule (12):

$$B_R := \{at(R, N1), B(O, at(R, N2)), \\ Past(at(R, N2)), B(O, Past(at(R, N2))), \\ G(R, at(R, N1)), G(O, at(R, N1)), \\ IK(O, G(R, at(R, N1))), \\ B(O, G(R, at(R, N1))), \\ B(O, IK(O, G(R, at(R, N1))), \\ B(O, G(O, at(R, N1)))\}$$

and proceeds to generate a response:

$$\Phi_{rev} := \{at(R, N1)\} \\ \Phi_{IK} := \{G(R, at(R, N1))\} \\ \Phi_{comm} := \{at(R, N1), G(R, at(R, N1))\} \\ \rightarrow Utt_0 := \{u_1, u_2, u_3, u_4\} \\ [[u_1]]_c := at(R, N1) \\ [[u_2]]_c := at(R, N1) \wedge Future(\neg at(R, N1)) \\ [[u_3]]_c := at(R, N1) \wedge G(R, at(R, N1)) \\ [[u_4]]_c := at(R, N1) \wedge Past(\neg at(R, N1)) \\ \rightarrow Utt_1 := \{u_1, u_3, u_4\} \\ \rightarrow NumRev([[u_1]]_c) := 1; NumRev([[u_3]]_c) := 2; \\ NumRev([[u_4]]_c) := 1 \\ \rightarrow Utt_{final} := u_3$$

Thus, R responds:

R: I am already at N1.

5 Discussion and Related Work

While the above case study was kept simple due to space restrictions, it demonstrates the utility of our utterance generation method in adapting NL output at the sentence-level based on a mental-model of an interlocutor. In particular, we adapted utterances by employing adverbial modifiers, which serve to make the speaker's belief-space more transparent and natural, which was the main motivation for the development of the formal framework with rules for adverbial modifiers in the first place. Other examples of adaptations that are intended to make an automated system's reasoning and internal state representations more open and clear to human-users include the sentence-level adaptation of restaurant recommendations (Walker et al., 2007) and the adapta-

tion of query-phrasing in a robotic context (Kruijff and Brenner, 2009). In addition to conveying information about one's own mental state, pragmatic principles and rules, such as those we have presented, may be deployed to reason about the intentions and beliefs of others (Perrault and Allen, 1980).

The current system, while a promising step towards more natural task-based dialogue interactions, has several limitations. Aside from lexical and semantic limitations, the currently implemented adverbial modifiers are restricted to very simple predicates. Clearly, these restrictions will have to be addressed and the formal definitions will have to be widened. Moreover, the system currently does not handle situations where a human's mental state changes without the robot's knowledge, which can cause misunderstandings that need to be detected and corrected effectively. Additionally, agents can be mistaken about their beliefs. Real-world complexities such as these suggest the inclusion of handling uncertainty in a belief modeling system (Lison et al., 2010), potentially by assigning beliefs confidence values. This is clearly an important topic for future work.

User-model based adaptation of NL output at the sentence level that includes multi-modal components (Walker et al., 2004) has also not been addressed. Further study is required to determine whether our Gricean-inspired utterance selection method can also be applied to non-linguistic communication modalities. Finally, the current system can only handle simple perceptual updates and has limitations when handling multi-robot dialogues (neither of which are discussed here for space reasons). The challenges of perceptual updates that will have to be addressed are investigated in the context of a plan-based situated dialogue system for robots in (Brenner, 2007) and extensions to multi-robot scenarios are explored in (Brenner and Kruijff-Korbyova, 2008).

6 Conclusion

Competency in mental modeling is a crucial component in the development of natural, human-like interaction capabilities for robots in mixed initiative settings. We showed that the ability to under-

stand and employ adverbial modifiers can help both in constructing mental models of human operators and conveying one's own mental state to others.

To this end, we made three contributions. First, we introduced a framework for formalizing different sentence types and the pragmatic meanings of adverbial modifiers. Second, we showed how one can perform belief updates based on implied meanings of adverbial modifiers. And third, we introduced a novel algorithm for generating effective responses that obey three Gricean maxims and aid the listener in appropriate belief updates. The core properties of the algorithm are that it corrects false or missing beliefs in other agents, that it provides an agent with information that is wanted, that it never generates an utterance that implies false propositions, and that it first favors utterances that convey more (true) propositions after favoring utterances that revise or add more beliefs to the listener's belief-space. Finally, we demonstrated our algorithm responding to basic operator queries in a simple case study, correctly using adverbial cues to sound more natural and convey more information regarding its beliefs.

There are extensive avenues to pursue future work. For instance, we plan to extend the algorithm to include multi-modal perceptual integration as well as multi-agent multi-dialogue capabilities. A variety of empirical evaluations would be desirable to evaluate the efficacy and naturalness of the proposed adverbial cues in simulated and real HRI tasks. Additionally, empirical evaluations could also be performed to observe additional cues to incorporate into the system.

7 Acknowledgments

This work was supported by an ONR MURI grant #N00014-07-1-1049 to the second author. We wish to extend our thanks to Paul Schermerhorn and the anonymous reviewers for providing valuable feedback.

References

- A. Kobsa. 2001. Generic User Modeling Systems. *User Modeling and User-Adapted Interaction* 11: 49–63.
- B. Di Eugenio, et al. 2008. Be Brief, And They Shall Learn: Generating Concise Language Feedback for a Computer Tutor. *International Journal of Artificial Intelligence in Education* 18(4).
- C. R. Perrault and J. F. Allen. 1980. A Plan-Based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics*, 6(3-4):167–182.
- G. M. Kruijff and M. Brenner. 2009. Phrasing Questions. *AAAI 2009 Spring Symposium*.
- H. Kress-Gazit and G. E. Fainekos and G. J. Pappas 2008 Translating Structured English to Robot Controllers *Advanced Robotics* 22, 12, 1343–1359
- H. P. Grice. 1975. Logic and conversation. *Syntax and Semantics*, 3(1):43–58.
- J. Dzifcak, M. Scheutz, C. Baral, and P. Schermerhorn. 2009. What to do and how to do it: Translating Natural Language Directives into Temporal and Dynamic Logic Representation for Goal Management and Action Execution. *ICRA*.
- M. A. Walker, et al. 2004. Generation and evaluation of user tailored responses in multimodal dialogue. *Cognitive Science* 28: 811–840.
- M. Walker, et al. 2007. Individual and Domain Adaptation in Sentence Planning for Dialogue. *Journal of Artificial Intelligence Research*. 30: 413–456.
- M. Scheutz, et al. 2007. First Steps toward Natural Human-Like HRI. *Autonomous Robots* 22(4):411–423.
- M. Brenner. 2007. Situation-Aware Interpretation, Planning and Execution of User Commands by Autonomous Robots. *RO-MAN 2007*.
- M. Brenner and I. Kruijff-Korbayova. 2008. Continual planning and acting in dynamic multiagent environments. *12th SEMDIAL Workshop*.
- P. R. Cohen, et al. 1990. Task-Oriented Dialogue as a Consequence of Joint Activity. *Pacific Rim International Conference on Artificial Intelligence*.
- P. Lison, C. Ehrler, and G. M. Kruijff. 2010. Belief Modelling for Situation Awareness in Human-Robot Interaction. *19th IEEE International Symposium*.
- R. Dale and E. Reiter. 1995. Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science*, 18(1):233–263.
- P. Rybski, K. Yoon, J. Stolarz, and M. Veloso. 2007 Interactive robot task training through dialog and demonstration *HRI*, 49–56
- T. Kollar and S. Tellex and D. Roy and N. Roy 2010 Toward Understanding Natural Language Directions *HRI*, 259-266
- W. Wahlster and A. Kobsa. 1989. User Models in Dialog Systems. *User Models in Dialog Systems*, 4–34. Springer-Verlag, Berlin.