The Pragmatic Social Robot: Toward Socially-Sensitive Utterance Generation in Human-Robot Interactions

Gordon Briggs  
NRC Postdoctoral Fellow  
Naval Research Laboratory  
Washington, DC 20375

Matthias Scheutz  
Human-Robot Interaction Laboratory  
Tufts University  
Medford, MA 02155

Abstract

One of the hallmarks of humans as social agents is the ability to adjust their language to the norms of the particular situational context. When necessary, they can be terse, direct, and task-oriented, and in other situations they can be more indirect and polite. For future robots to truly earn the label “social,” it is necessary to develop mechanisms to enable robots with NL capabilities to adjust their language in similar ways. In this paper, we highlight the various dimensions involved in this challenge, and discuss how socially-sensitive natural-language generation can be implemented in a cognitive, robotic architecture.

Introduction

As technology enabling more sophisticated and robust autonomous agents continues to advance, there is growing interest in the development of so-called “social robots.” These are robotic agents that are designed and marketed as having increasingly natural and/or human-like “social” interaction capabilities. In particular, there is increasing interest in robotic agents that have natural language (NL) communication capabilities. Yet, simply possessing basic NL capabilities is not sufficient to reproduce the rich social facets of human NL communication. To create truly “social” robots, richer NL mechanisms are needed.

Specifically, one of the key aspects of human-level NL communication is the fact that people are able to use a variety of different linguistic forms (both literal and non-literal) to communicate a single intention. For instance, someone could request a cup of coffee in a combinatorially explosive number of ways: “Get me a coffee, now!”; “Get me a coffee.”; “Get me a coffee, please.”; “Would it be possible for me to have a coffee?”; “A coffee would be great...”, etc.

How people are able to resolve this multitude of different forms to a particular intention is the domain of pragmatics. Indeed, there is a set of prior work focused on enabling computational agents to understand indirect speech acts (ISAs) (Perrault and Allen 1980; Wilske and Kruijff 2006; Briggs and Scheutz 2013; Williams et al. 2015). However, there still exists the reverse problem: how do we enable our putative “social” robots to reciprocate and display this rich diversity of language. How do enable them to select the most appropriate linguistic form given the current social context?

In this paper, we highlight an underexplored aspect of socio-linguistic behavior in HRI and AI, specifically the modulation of generated NL based on pragmatic and socio-linguistic factors. We first briefly introduce a set of social and pragmatic goals that are often considered when realizing speech acts. We then present the utterance selection mechanism utilized in the DIARC/ADE robotic architecture that factors in some of these pragmatic goals. Finally, we discuss the limitations of our current architecture and future areas of development and research.

Pragmatic and Socio-Linguistic Goals

People balance a variety of communicative goals when formulating an utterance. These include criteria found in Grice’s conversational maxims (Grice 1975) and other normative social criteria such as politeness (Brown and Levinson 1987; Leech 2014). In this section, we present a series of possible communicative goals that should be considered when formulating and selecting utterances in HRI.

Politeness

There has been a growing interest in investigating robotic or virtual agents that utilize politeness when making requests of human interactants (Gupta, Walker, and Romano 2007; Torrey, Fussell, and Kiesler 2013; Strait, Cannig, and Scheutz 2014; Srinivasan and Takayama 2016). Less common, but still present, is work devoted to the development of computational mechanisms to appropriately implement these strategies (Briggs and Scheutz 2013; Gupta, Walker, and Romano 2007; Miller et al. 2007). These previous studies, however, contain some limitations.

The first limitation is a narrow focus on what Brown and Levinson (1987) deem negative politeness, which pertains to the degree to which an utterance constrains the subsequent action of the addressee. Specifically, most of these prior works focus on indirectness as the primary means by which to vary the politeness of a directive. Indirect commands, such as “Could you do X?”, are considered more polite than direct forms as they do no force the addressee to explicitly accept or reject a potentially obligated action,
rather it simply triggers a conversational obligation to inform the speaker as to one’s ability. However, the focus on directness often leads to seemingly counterintuitive findings in which highly indirect request forms are evaluated to be most impolite (Gupta, Walker, and Romano 2007; Srinivasan and Takayama 2016). This is usual due to the fact that many of the highly indirect forms used in these studies (e.g. “You haven’t done the job yet”) often sounding passive-aggressive or accusatory. Future work needs to be done to model threats to positive politeness, which pertains to the degree to which an utterance damages the image, feelings, or reputation of the addressee. We have highlighted this limitation in (Briggs and Scheutz 2014).

The second limitation is the conflation of what Leech (2014) deems the pragmalinguistic and the sociopragmatic aspects of politeness. The pragmalinguistic aspect of politeness is context independent (e.g. directives given in the form of questions are more polite than direct commands, or saying “please” is more polite than not saying “please”), whereas the sociopragmatic aspect pertains to the appropriateness of an utterance in the present social context (i.e. social relationship and distance between the two interlocutors). Most of the HRI/HCI studies on politeness in requests has adopted a primarily pragmalinguistic stance, in which politeness modulation is achieved by the application of a particular linguistic politeness strategy, but no explicit reasoning about whether or not the directive is appropriate given the particular social relationship between the speaker and the addressee. For example, it would be inappropriate for a janitor to direct his or her supervisor to sweep the floor for him or her, regardless of whether it was phrased as an ISA or modifiers such as “please” were used.

**Brevity and Clarity**

While politeness is a major sociopragmatic goal, it often comes into conflict with other pragmatic goals. One aspect of Grice’s Maxim of Manner states that a speaker should “Be brief!” and to be clear (Grice 1975). Certain social contexts, such as critical task-based scenarios, necessitate efficient and unambiguous communication. The use of indirect speech to satisfy politeness considerations in these scenarios would be inappropriate (for example, image a neurosurgeon asking an assistant in the midst of an operation, “Would you mind handing me a scalpel, please?”).

**Informativeness**

Another of Grice’s conversational maxims states that a communicative contribution should be as informative as possible (the Maxim of Quantity) (Grice 1975). People often use subtle cues to convey additional information. For instance, consider the subtle difference between the following sentences:

1. I’m not at the store.
2. I’m not at the store, yet.

The first statement conveys the simple fact that the speaker is not currently at the store. The second statement conveys this fact while additionally conveying that the speaker as an expectation that he or she will be at the store at a later time. It is equivalent to saying, “I’m not at the store, but I plan to be at some point.” Indeed, this adverbial modifier is a way of conveying this additional information while also trying to satisfy the pragmatic goals of brevity!

**Pragmatics Framework**

Having presented a set of pragmatic and social criteria, we now present the mechanisms used in the DIARC/ADE robotic architecture (Scheutz et al. 2013) to select a particular utterance form given a particular intention to convey in ways sensitive to these criteria. In the architecture, the following utterance representations is used:

\[ U = \text{UtteranceType}(\alpha, \beta, X, M) \]

where UtteranceType denotes the speech act classification, \( \alpha \) denotes the speaker, \( \beta \) denotes the addressee, \( X \) denotes an initial semantic analysis, while \( M \) denotes a set of sentential modifiers (e.g., “now”, “still”, “really”, “please”). These utterance type representation then form the basis for representations of dialogue exchange patterns and pragmatic inference rules.

**Pragmatic Rules**

There are two main functions of the pragmatic rule system within DIARC/ADE’s dialogue component. First, it determines a set of beliefs (B) to infer upon receiving a natural language utterance \( U \) in a given context \( C \). Second, it can also determine what the best and most appropriate utterance form \( U \) is to generate given the goal of communicating the formula \( \phi \) in context \( C \), which is the focus of this paper.

We employ a representational form of pragmatic rules described below, which we first present in (Briggs and Scheutz 2013):

\[ [[U]]_C := (B_{lit}, B_{int}, \theta) \]

The rule associates a particular utterance form \( U \) in context \( C \) with a tuple containing the set of beliefs \( B_{int} \) to be inferred based on the intended meaning of the utterance (which is potentially non-literal), the set of beliefs to be inferred based on the literal meaning of the utterance \( B_{lit} \), as well as the degree \( \theta \) to which the utterance can be considered a face-threatening act (FTA) in context \( C \). The context \( C \) is currently represented as a set of facts that must obtain in order for the pragmatic rule to be considered. The belief component in DIARC/ADE, which contains the knowledge base of the robot, is constantly queried to maintain the subset of applicable pragmatic rules at a given time.

The notion of an FTA is drawn from politeness theory (Brown and Levinson 1987). Operationalizing the degree to which an utterance can be construed as a FTA in a given context has been used in previous systems (Gupta, Walker, and Romano 2007; Miller, Wu, and Funk 2008; Briggs and Scheutz 2013). Both belief sets are represented in order to determine whether or not the pragmatic rule corresponds to a literal form, which affects the modulation of natural language generation based on politeness considerations.
Utterance Selection

We view utterance selection in HRI as a two-step process: (1) determining what to say as part of a larger agent plan, (2) selecting a particular utterance form to realize the planned speech acts. Most work in dialogue-enabled agents focuses on the challenges in the first step, which is sufficient to enable basic NL interaction. However, we believe consideration of both steps is necessary to enable truly natural and “social” NL interaction. Below we discuss how social factors can be incorporated at both stages of natural language generation.

What To Say In human-initiative dialogue interactions, the robot can respond to a human directive or query reactively by communicating a belief or intention φ, which represents the relevant answer or acknowledgment of the human’s directive. However, in a robot-initiative dialogue interaction, the robot must formulate a plan to achieve a goal which may include sequence of speech actions. To explicitly model agents deciding to adopt another agent’s goal as their own in a planning context (in response to a request for instance), Perrault and Allen (1980) present a “glue” actions. These glue actions can be tailored such that appropriate social role or ranking can be a precondition for these actions to apply (Briggs and Scheutz 2013). This leads to sociopragmatic modulation of speech at the planning stage.

How To Say It Given a desired, STRIPS-style, speech action A to realize (such as those defined in (Perrault and Allen 1980)), we can determine an associated predicate to convey φ, or φ can be supplied directly. Given a speech action A and its set of effects A-effects, and a pragmatic rule with a set of intended belief updates B-int, we consider it plausible that the action A and the utterance U in context C associated with the pragmatic rule are equivalent if the following holds:

\[
[[U];_C := (B_{int}, B_{int}, \theta)) \land \\
([\phi \in B_{int} \land \phi \in A_{effects}) \lor \\
(want(\alpha, bel(\beta, \phi)) \in B_{int} \land bel(\beta, \phi) \in A_{effects}) \lor \\
(\phi \in B_{int} \land \phi \in A_{effects}))
\]

For instance, consider the question, “Can I have a coffee?”, which is represented as AskYN(S, H, capableOf(S, have(S, coffee)), {}) in utterance form. If in a particular context this is plausibly interpreted as an indirect request with meaning B-int = \{want(S, do(H, have(S, coffee)))\}, then there are (at least) two possible ways of realizing the request speech action request(S, H, have(S, coffee)):

1. Instruct(S, H, do(H, have(S, coffee)), \{}): As φ = want(S, do(H, have(S, coffee))) and \phi \in B-int \land bel(H, \phi) \in A-effects.
2. AskYN(S, H, capableOf(S, have(S, coffee)), \{}): As φ = want(S, do(H, have(S, coffee))) and \phi \in B-int \land bel(H, \phi) \in A-effects.

Then, given the desired piece of information to convey, we can to find a set of pragmatic rules R_A where \phi \in B-int for all the rules R \in R_A. Recall that B-int is the set of inferred belief updates resulting from the application of the pragmatic rule, which may be non-literal. We then take the subset of rules where \theta \leq \Theta_{max}(C), for a contextually determined maximum face threat (FT) threshold, where \theta is the degree to which the utterance can be construed as a FT2. Then the remaining rules are sorted using the following criteria (in order of descending priority):

1. Favoring rules with the lowest number of false or unsupportable predicates in B-int (Grice’s Maxim of Quality (Grice 1975)).
2. Favoring rules that are direct, such that B-int = B-int. (Grice’s Maxim of Manner).
3. Favoring rules with the highest number of predicates that correct false beliefs of the addressee or are novel to the addressee. (Grice’s Maxim of Quantity).
4. Favoring rules with the highest number of supportable predicates. (Grice’s Maxim of Quantity).
5. Favoring rules with the lowest \theta values. (Minimize face-threat potential all other things being equal).

The first (highest-ranked) rule from the sorted list is then selected and sent to the downstream components in the natural-language generation (NLG) pipeline. This approach is an extension of the utterance selection approaches found in (Briggs and Scheutz 2011; 2013).

The ordering of the rules was designed to conform with intuitions regarding the relative prioritization of social/communicative norms. Making sure that one is not saying something that could be construed as lying (saying that for which the negation is believed) or confabulating (saying that for which there is no evidence/support either way), is of top priority. After that, being clear and direct is favored (as utterances with an inappropriate FT-degree value have already been filtered out at this point), followed by considerations of being as informative as possible. Finally, politeness considerations can be used to break ties.

Discussion

Having presented the general approach to utterance selection in the DIARC/ADE robotic architecture, we can highlight the contributions and limitations of this approach. The mechanisms that we have described have been used in a variety of prior work to begin to achieve the various pragmatic goal desirata specified above. For instance, the utterance selection/ranking system balances concerns of clarity and politeness by favoring direct forms in contexts in which the FT-threshold is high. Informativeness is considered, and has been demonstrated in prior work on appropriate usage of adverbal modifiers (Briggs and Scheutz 2011). Finally, unlike most of the prior work on politeness in HCI/HRI, our system has demonstrated an explicit consideration of sociopragmatic politeness, rather than just pragmalinguistic politeness (Briggs and Scheutz 2013). Yet, while our utterance selection mechanism has pushed the bounds of what previous systems are capable of, there is still a lot of future work to be done to realize the ultimate goal of sophisticated social speech.

---

2 Note, we use the terms face threat (FT) and face threatening act (FTA) interchangeably in this section.
Limitations and Future Research

One of the major current limitations of the present system is that it utilizes a fixed strategy that is hand-crafted to yield appropriate results. Ideally, the utterance ranking and selection mechanism should determine different rankings of candidate utterances given a specific pragmatic goal and how to select an appropriate response given the subset of communicative goals activated in the current social context. However, to realize this more flexible pragmatic generation system, a large amount of future research needs to be accomplished, which we discuss below.

Brevity Currently, candidate utterances are not ranked by how brief their surface realizations may be. This requires future integration of the pragmatic rule framework with lower-level NLG and text-to-speech components. It also requires a particular operationalization of what brevity is. Is it the time it takes for the text-to-speech component to speak the utterance? Is it the number of words or phonemes in the text associated with the utterance? This is certainly a question for future investigation.

Politeness As discussed previously, politeness is a complex linguistic and social phenomena. Too much directness is not the only way in which utterances can be impolite. As such, continued work is necessary to expand politeness evaluation systems to predict and model phenomena such as positive politeness (Briggs and Scheutz 2014).

Balancing Goals It is usually not the case that a single pragmatic goal is active, rather it is a matter of determining what the current preference ordering of communicative goals is in given situation (which is in and of itself another future research challenge). We have recently developed a more flexible utterance selection mechanism that selects utterances based on a preference ordering communicative goals. However learning what the appropriate preference ordering of these communicative goals are in different HRI contexts remains to be explored.

Conclusion

One of the hallmarks of humans as social agents is the ability to adjust their language to the norms of the particular social context. When necessary, they can be terse, direct, and task-oriented, and in other situations they can be more indirect and polite. For future robots to truly earn the label “social,” it is necessary to develop mechanisms to enable robots with NL capabilities to adjust their language in similar ways. In this paper, we have highlighted the various dimensions involved in this challenge, and presented the first steps towards enabling socially-sensitive in the DIARC/AR cognitive, robotic architecture.

References


