What Social Robots Can and Should Do
J. Seibt et al. (Eds.)
IOS Press, 2016

doi:10.3233/978-1-61499-708-5-293

Polite Interactions with Robots

Luciana BENOTTI [a,1] and Patrick BLACKBURN [b]
[a] *Section of Computer Science, Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Argentina*
[b] *Section of Philosophy and Science Studies, Centre for Communication and Arts, University of Roskilde, Denmark*

Abstract. We sketch an inference architecture that permits linguistic aspects of politeness to be interpreted; we do so by applying the ideas of politeness theory to the SCARE corpus of task-oriented dialogues, a type of dialogue of particular relevance to robotics. The fragment of the SCARE corpus we analyzed contains 77 uses of politeness strategies: our inference architecture covers 58 of them using classical AI planning techniques; the remainder require other forms of means-ends inference. So by the end of the paper we will have discussed in some detail *how* to interpret automatically different forms of politeness — but *should* we do so? We conclude with some brief remarks on the issues involved.

Keywords. politeness theory, positive face, negative face, face threatening acts, situated interaction, task-oriented dialogues, conversational implicature

1. Introduction

Should humans be polite to robots? And should robots be polite to humans? Before addressing such issues, it seems interesting to consider what politeness is, and what it takes to program robots with the ability to converse politely. Brown and Levinson [1] have given an influential account of how politeness is embodied in language; in this paper we show how their account can be handled computationally. We do so by building on our previous work on inferring conversational implicatures in task-oriented dialogues.

We have previously [2,3,4] argued that Grice's notion of conversational implicature is best viewed as an embodied method of negotiating meaning. 'Embodied' because — as Grice [5] himself emphasized — speech acts are genuine acts, and both physical and linguistic acts can be involved in determining meaning. 'Negotiated' because embodied agents operate under uncertainty — and iterated acts of mutual recognition and correction lie at the the heart of human communication. We showed that many of the inferences involved in such interactions can be modeled using classical AI planning [6] and implemented using off-the-shelf AI planners; we also argued that forms of means-ends inference beyond the scope of classical planning are sometimes required.

Politeness fits into this Gricean picture. Politeness (like implicature) is a pragmatic phenomenon involving interaction and inference, but it brings a new ingredient into play:

[1] Corresponding Author: Luciana Benotti, Section of Computer Science, Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Argentina; E-mail: benotti@famaf.unc.edu.ar.

face. However, as in our earlier work, AI planning turns out to model the core inferences involved in polite interactions, though (as with implicature) extended forms of means-ends inference extending this paradigm are sometimes required.

In this paper we briefly discuss the linguistic theory we shall use, indicate the politeness phenomena that interest us by annotating a human-human corpus, and sketch an inference architecture that is able to understand them. With this done, we briefly note what our work suggests about the type of *should* questions with which the paper starts.

2. Politeness Theory

Politeness theory was first formulated by Brown and Levinson [1]. The theory accounts for the strategies that speakers use in order to avoid damaging their own face, or threatening their hearer's; here face is used in the sense of social prestige or self image. A 'face threatening act' is an act that may damage the speaker's or hearer's face. Politeness theory assumes that most speech acts inherently threaten one or both of these faces, and views politeness as an essential component of *non*-face-threatening communication.

There are two aspects to face, positive and negative: 'positive face' refers to one's self-esteem, 'negative face' refers to one's freedom to act. These two aspects of face are considered to be basic wants in any social interaction, and politeness strategies are seen as soothing layers that are applied to face threatening acts to ensure that face is preserved. The greater the potential for loss of face, the stronger the politeness strategy that should be used.

Politeness theory distinguishes between four different politeness strategies: 'bald on record', 'positive politeness', 'negative politeness' and 'off the record'. In the following sections we illustrate each of the strategies and discuss the knowledge representation and inference mechanisms needed in order to understand them.

3. Looking for Politeness Phenomena in Human-Human Interactions

We illustrate the different politeness strategies with examples from the the SCARE corpus [7], a human-human corpus of interactions consisting of fifteen spontaneous English dialogues associated with an instruction giving task.[2] As in most situated dialogues, in the SCARE corpus the language used is deeply grounded in the game world the dialogue participants share. The corpus was collected using the QUAKE environment, a first-person virtual reality game.

The task consists of a direction giver (DG) instructing a direction follower (DF) on how to complete several tasks in a simulated game world. The DF has no prior knowledge of the world map or tasks and relies on his partner, the DG, to guide him in completing the tasks. The partners speak to each other through headset microphones; they cannot see each other. As the participants collaborate on the tasks, the DG has instant feedback of the DF's location in the simulated world, because the game engine displays the DF's first person view of the world on both the DG's and DF's computer monitors. We selected this corpus because modeling the role of the DF (who needs to be able to interpret polite instructions and follow instructions politely) poses challenging problems for robotics.

[2] The corpus is freely available for research at http://slate.cse.ohio-state.edu/quake-corpora/scare/.

In Figure 1 we show the percentage of the four different politeness strategies that we found in 125 face threatening acts that we annotated from the SCARE corpus. We present examples from the SCARE corpus for each of them in this section.

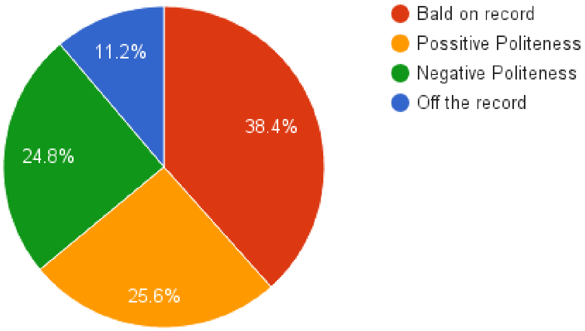

Figure 1. Percentage of the four different politeness strategies used in the face threatening acts we annotated in the SCARE corpus (n=125).

3.1. Bald on Record

When no soothing layer is put on the act, the speaker is performing 'bald on record'. Such face threatening acts are usually realized as imperatives. Below we give examples of bald on record face threatening acts found in the SCARE corpus.

- *Out through that door.*
- *Grab that thing.*
- *Keep going straight through.*
- *Let go of that thingy.*
- *Press the left button, the one to your left.*

The SCARE corpus contains a large number of bald on record acts. In the fragment of the corpus that we annotated, 34.8% of the face threatening acts were performed bald on record. There may be a number of reasons for this. To begin with, the dialogue participants were friends; there existed a certain confidence between them. The tasks were long and sometimes repetitive. The dialogue participants were performing the experiment under time pressure. Bald on record realizations of requests are a short and effective way of communicating and these advantages sometimes trump the need to preserve face.

According to politeness theory, speakers choose the strategy needed for a face-threatening act by considering the combination of three variables: power, distance, and rank. 'Power' refers to the perceived hierarchical relation between speaker and hearer. That is, the power variable can be set by asking: Is the targeted hearer a superior, a subordinate, or at about the same social level as the speaker? In the case of the SCARE corpus, most experimentees were college students, and may be taken to belong to roughly the same social level. 'Distance' refers to the degree of social distance between speaker and hearer. This variable answers the question: Is the targeted hearer a close friend or a distant colleague of the speaker? In the case of the SCARE corpus, the hearer was always a friend. 'Rank' refers to the cultural ranking of the topic being talked about — the degree of sensitivity of the topic within a particular culture. For example, a woman's age and weight are two highly sensitive topics within U.S. culture, as is a person's income in Danish culture; but many other cultures don't consider these sensitive topics, they simply regard them as matters of fact to be shared. In the SCARE corpus, the topic was not a sensitive one: they were just talking about a fictional task on a game world. That is, they were talking while playing a video game together.

3.2. Positive Politeness

In positive politeness, the speaker's goal is to address the positive face needs of the hearer: that is, the goal is to enhance the hearer's positive face (this is also known as 'positive face redress'). Positive politeness strategies highlight friendliness and camaraderie between the speaker and hearer; the speaker's wants are taken to be in some way similar to the hearer's wants. There are many ways to accomplish this familiarity and thus claim common ground.

First, the speaker can notice and attend to the hearer's wants, interests, needs, or goods. The speaker can also use in-group markers, which demonstrate that both the speaker and hearer belong to the same social group. These can include altering the forms of address or using in-group language. For instance, the use of "we" instead of "you", as in the example below, is a kind of in-group marker. Other kinds of in-group language include the use of dialects, jargon or slang, and linguistic contractions. Here is an example of the linguistic contraction "wanna" being used in this way:

Example PP1: Claim in-group membership via inclusive forms of address.
DG(1): we wanna move those boxes so they're all three on the left side of the table
DF(2): ok may be I try jumping in and up there
DG(3): I'm not sure . uh . may be you can just press control .
DF(4): I tried that and no luck

The speaker can also seek agreement with the hearer by choosing safe topics and using repetition. Or a speaker may seek to avoid disagreement by employing a token agreement, a confirmation question, or a white lie.

Example PP2: Seek to avoid disagreement via confirmation questions.
DG(1): see that button straight ahead of you?
DF(2): mhm
DG(3): hit that one

The DG does not directly say "hit the button straight ahead of you" because he does not want to assume that the DF sees the button (for if the DF has not, this threatens his positive face). Instead the DG asks. This causes the DF to focus his attention. Once this is achieved, the DG says: hit that one.

Finally, in an effort to establish positive politeness, the speaker can claim common ground with the hearer in some way. For example, this can be done by presupposing the hearer's knowledge of the topic, or by seeking cooperation through overstatement.

Example PP3: Claim common ground and seek cooperation by overstating.
DG: go all the way back through ... you know ... we've done this like a million times

Obviously, the DF has not done this task a million times before, but the DG is claiming common ground with the DF via these utterances. In response, the DF laughs and starts following the instruction. Positive face was preserved.

3.3. Negative Politeness

Whereas positive politeness enhances the hearer's positive self-image via recognition of the hearer's need for his or her wishes and desires to be appreciated socially, negative politeness addresses the hearer's need for freedom of action, and freedom from imposition in decision making (this is also known as 'negative face redress').

This strategy seeks to minimize the imposition on the hearer. The first approach to negative politeness is not to assume or presume. A useful strategy here is the use of hedges. A hedge is a softening of a statement, induced by employing less-than-certain phrasing such as might, could, should, etc.

Example NP1: Use hedges for less-than-certain phrasing.
DG: you should be able to continue the way you came in

A second approach is to not coerce the hearer. This can be accomplished through apologetic language, such as giving reasons for having to ask.

Example NP2: Avoid coercing the hearer by giving reasons for having to ask.
DG: you have to move the boxes to a certain configuration, there is a button that does it, its the left one
DF: this one?
DG: yeah, go for it

Further efforts to avoid encroaching on the hearer's sense of freedom include impersonalizing the speaker and hearer. Common strategies here include using passive and circumstantial voices ("It's generally done this way"), replacing "I" and "you" with indefinites, and avoiding use of "I" and "you" all together.

Example NP3: Impersonalize speaker and hearer by avoiding their pronouns.
DG: alright, the middle box, it needs to be moved forward
DF: ok, let me see if I can

In the example the DG says "it needs to be moved forward" avoiding the use of "I" and "you". Of course, the only way to actually move the box forward is for the DF to perform the task.

3.4. Off the Record

Off-the-record is a politeness strategy that relies upon the speaker dropping hints and relying on the hearer to figure out what is intended. This strategy is highly indirect, and in the setting of the SCARE corpus involves the breaking of conversational norms to implicitly suggest that a particular course of action is recommended; the speaker relies on the hearer's ability to decipher and interpret the intended meaning. In essence, this strategy boils down to being ambiguous in a way that allows the speaker to avoid claiming responsibility for choosing the act. Off-record politeness can be accomplished in a various ways, but they typically involve the speaker inviting Gricean conversational implicatures [5], thereby letting the hearer take the initiative and thus avoiding imposing a restricted course of action on him.

In the following example the DG says in (1) that a cabinet is "back where they started". The DG would not have said so if interacting with this cabinet were not relevant at this point in the task. Hence the DG implicates that the DF at some point will need to go back to where he started. The DF takes this up, confirming each implicated step with a clarification request in turns (2), (4) and (6). (Note that the DF is using a positive politeness strategy to follow the implicated instructions: seeking agreement and reaffirming the common ground with the DG.)

Example OR1: Invite conversational implicatures (maxim of relevance).[3]
DG(1): it's . kinda like back where you started . so
DF(2): ok . so I have to go back through here?
DG(3): yeah
DF(4): and around the corner?
DG(5): right
DF(6): and then do I have to go back up the steps?
DG(7): yeah
DF(8): alright, this is where we started

Secondly, the speaker can be intentionally vague or ambiguous, leaving longer silences than standard ones in a conversation, and being incomplete by using ellipsis. In the following example, the DG and the DF both observed that a cabinet opened as a result of pressing the button on the left. The DG did not issue an instruction after this event. This silence gave the DF the opportunity to take the initiative by inferring the next relevant step of the task in turn (5).

Example OR2: Invite conversational implicatures (maxim of quantity).[4]
DG(1): press the button on the left [pause]
DG(2): and . uh [pause]
DF(3): [pause]
DG(4): [pause]
DF(5): put it in this cabinet?
DG(6): put it in that cabinet, yeah

In the following exchange the DG is violating the maxim of quality because he knows that the DF cannot answer question (1). Only the DG has this knowledge. The DF also knows this, and thus interprets (1) as a rhetorical question and gives the DG time to answer himself in (3).

[3] Grice's maxim of relevance [5] is: Be relevant.

[4] Grice's maxim of quantity [5] has two parts: (1) Make your contribution as informative as is required (for the current purposes of the exchange) and (2) Do not make your contribution more informative than is required. Though it could be argued that it is Grice's maxim of manner (be perspicuous) that is being violated here.

Example OR3: Invite conversational implicatures (the maxim of quality).[5]
DG(1): now, what are we doing here?
DF(2): [pause]
DG(3): there . you should find a chair in a corner

4. Computing Polite Abilities for Robots through Automatic Inference

In this section we sketch an inference architecture capable of undestanding the politeness phenomena described and illustrated in Section 3. We indicate which parts fall within the classical AI planning paradigm, and which require more. Table 1 summarizes our results.

4.1. Positive Politeness

In order to interpret the instruction "we wanna move those three boxes so they're all three on the left side of the table", the DF can use classical AI planning by indicating that the DG is one of the agents available for performing certain actions. The use of this in-group marker might mean that the DG is available if his help is needed (for example, to provide further information and guidance) although he cannot act in the game world. As is evident from his responses to the instruction, the DF inferred alternative plans, silently executed one of them (pressing control), and is looking for approval from the DG on the second one (jumping on the table). Classical AI Planners such (as [8]) can return more than one plan for a given goal; a minor modification to their code enables them to keep looking when they find a first plan.

PP2 cannot be handled by using just classical AI planning. PP2 is a way of putting facts from the world in focus, thereby activating the preconditions of the following acts. But classical AI planning does not provide a way to put fact in focus: facts are either true or false for AI planners. As a result, although the instruction (1) from example PP2 can be interpreted, its effect on the elements of the world that are in focus, and which are needed in order to interpret "that button" in (3), requires representation and inference machinery that lies outside the classical AI planning paradigm.

PP3 involves interpreting instructions based on the previous story of the interaction. If the story of the interaction is represented as part of the common ground, a planner can be used to interpret the instruction of going-back-again-to-some-position in the game world; it would do so by taking this common ground representation as its initial state when interpreting the instruction in Example PP3.

4.2. Negative Politeness

NP1 is similar to PP3: a plan has to be searched for it using the common ground representation as the initial state of the planning problem. The instruction includes hedges such as 'should' or 'I guess' that can be ignored during its interpretation if a plan is found in the common ground.

NP2 explains the plan and its effects in a top down fashion in order to let the DF know the reasons for the requested actions. The authors of [9] argue that this is a strategy that improves agreement and diminishes misunderstandings in task oriented interactions

[5] In its most general form Grice's maxim of quantity [5] is: Try to make your contribution one that is true.

such as the one in SCARE. This cannot be modeled with classical AI planners where the plans returned are flat (that is, without hierarchical structure). Some sort of hierarchical planner would be needed here.

NP3 is an strategy where personal pronouns are avoided. As in PP1, an AI planner can be used to infer which agents are involved in achieving the goal stated by the face threatening act. However this negative politeness strategy is sometimes realized not only by omitting the agent but also the verb. For instance, the DG said "the door on the right" and the DF had to infer that he himself had to go through that door. This can be inferred unambiguously when the object does not afford other actions and it is clearly a form of means-end reasoning. However, such affordances of objects can not be obtained using just classical AI planners.

4.3. *Off the Record*

Example OR1 is an off-the-record way of asking to go back where they started that can be interpreted using AI classical planning; we show this in detail in [2].

The inference that the DF did in Example OR2 in order to produce (5) can be defined as another means-end inference task which involves finding the 'next relevant actions'. The input of such task would consist of an initial state, a set of possible actions, and will also contain one observed action (in the example, action (1)). Inferring the next relevant action consists in inferring the set of executable actions of the initial state and the set of executable actions of the state after the observed action was executed. The 'next relevant actions' will be those actions that were activated by the observed action. In the example above, the next relevant action that will be inferred is "put the thing you are carrying in the cabinet that just opened", just what the DF predicted in (5).

OR3 starts with a question. Instructions that clearly cannot be followed by the DF or questions that cannot be answered by him are typically interpreted as rhetorical questions. AI planners can be used to interpret questions as rhetorical: simply wait for the DG to answer himself after realizing that no plan can be found.

In Table 1 we show the frequencies and classification of the three politeness strategies implemented by the face threatening acts that we annotated from the SCARE corpus (n=125). 48 acts are realized as bald on record and thus implement no politeness strategy. The other 77 acts are classified following the categories that we illustrated in Section 3. In column 'Can be inferred with classical AI planning' we indicate which of them fall within the classical AI planning paradigm.

5. Looking Ahead

In this paper we sketched an inference architecture for interpreting aspects of politeness as embodied in natural language. Our treatment was restricted to politeness in task oriented dialogues, but these are dialogues of clear relevance to contemporary robotics. We showed that 58 of the 77 inferences in the fragment of the SCARE corpus we annotated can be handled in the classical AI planning paradigm (which means they can be implemented using off-the-shelf planners). The other 19 require other forms of inference; some of these are likely to be covered by new-generation software tools, others require further research.

Politeness strategy	**Occurrences in the SCARE corpus (n=125)**	**Can be inferred with classical AI planning**
PP1 in-group language	17	17
PP2 confirmation questions	7	0
PP3 claim common ground	8	8
NP1 less-than-certain phrasing	13	13
NP2 giving reasons for asking	6	0
NP3 Impersonalize the speaker and hearer	12	8
OR1 maxim relation	8	8
OR2 maxim quantity	2	0
OR3 maxim quality	4	4
Total	77	58

Table 1. Frequencies and classification of the three politeness strategies implemented by the face threatening acts that we annotated from the SCARE corpus (n=125). 48 acts are realized as bald on record and thus implement no politeness strategy.

But what of the two questions at the start of the paper: Should humans be polite to robots? And should robots be polite to humans? Such questions take us from ‘how to program it’ to ‘whether we should’ — does our work offer any insight on this issue? We believe it does.

Here is one reason. Our approach makes heavy use of the ideas of Grice. In essence, Grice’s classic work on implicature treats language use as situated interaction. In order to handle real world dialogues, linguistic agents require far more than knowledge of language (in the sense of phonological, syntactic and semantic competency). They also need the ability to reason about other agent’s intentions, beliefs and desires, and to relate these to the norms that cover conversation (and to spot possible violations). The concepts of politeness theory further increases the complexity of the situations in which agents must work, most notably by insisting on the omnipresence and importance of positive and negative face.

In what sort of situations are such styles of reasoning necessary? Here it is helpful to use Dennett’s terminology: we are looking at situations where taking the ‘intentional stance’ [10] is useful. But until comparatively recently, taking the intentional stance (at least for language-using agents) was something done by human beings when thinking about other human beings. The analysis we have given — and indeed, vast amounts of work in contemporary robotics and AI — indicates that the machines we create (which until recently could be adequately analyzed using what Dennett calls the *design stance*) increasingly need to be viewed via the intentional stance. Furthermore, it suggests such machines will need to adopt it themselves in their own interactions with other agents, whether machine or human. The once-clear line between agents for which the intentional and design stances were appropriate is becoming increasingly blurred; indeed we might say that a major goal of both robotics and AI is to blur them ever more thoroughly.

Contemporary ethical discussion draws heavily on the concept of intentions, beliefs and desires. How will these concepts fare in a complex world of complex agents (human, human-made and machine-made) which need to be thought of, and are capable of thinking in terms of, various gradations of the intentional stance? No doubt practical answers can be given for simple robots, but these don’t begin to exhaust the scale of the complexity we shall soon be facing. Interesting times are ahead.

References

[1] P. Brown and S. Levinson. *Politeness: Some universals in language usage*. Studies in Interactional Sociolinguistics, 1978.

[2] L. Benotti. *Implicature as an Interactive Process*. PhD thesis, Université Henri Poincaré, INRIA Nancy Grand Est, France, supervised by P. Blackburn, 2010.

[3] L. Benotti and P. Blackburn. Classical planning and causal implicatures. In M. Beigl, H. Christiansen, T.R. Roth-Berghofer, A. Kofod-Petersen, K.R. Coventry, and H.R. Schmidtke, editors, *Modeling and Using Context*, volume 6967 of *Lecture Notes in Computer Science*, pages 26–39. Springer Berlin Heidelberg, 2011.

[4] L. Benotti and P. Blackburn. Context and implicature. In P. Brézillon and J.A. Gonzalez, editors, *Context in Computing: A Cross-Disciplinary Approach for Modeling the Real World*, pages 419–436. Springer New York, New York, NY, 2014.

[5] P. Grice. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics*, volume 3, pages 41–58. Academic Press, New York, 1975.

[6] D. Nau, M. Ghallab, and P. Traverso. *Automated Planning: Theory & Practice*. Morgan Kaufmann Publishers Inc., CA, USA, 2004.

[7] L. Stoia, D.M. Shockley, D.K. Byron, and E. Fosler-Lussier. SCARE: A situated corpus with annotated referring expressions. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakesh, Morocco, May 2008.

[8] M. Helmert. The fast downward planning system. *Journal of Artificial Intelligence Research*, 26(1):191–246, July 2006.

[9] M.E. Foster, M. Giuliani, A. Isard, C. Matheson, J. Oberlander, and A. Knoll. Evaluating description and reference strategies in a cooperative human-robot dialogue system. In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, IJCAI'09, pages 1818–1823, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[10] D. Dennett. *The Intentional Stance*. MIT presss, 1996.