Service Robots dealing with indirect speech acts

Sabrina Wilske & Geert-Jan Kruijff
Language Technology Lab
German Research Center for Artificial Intelligence (DFKI)
Saarbrucken, Germany
Email: {wilske, gj}@dfki.de

Service Robots dealing with indirect speech acts

***Abstract*— Successful interaction between a service robot and its human users depends on the robot's ability to understand not only direct commands, but also more indirect ways for a human to express what she would like the robot to do. Such indirect ways are pervasive in human-human interaction; enabling the robot to understand them can make human-robot interaction more human-friendly. This paper presents a model for a robot that pursues its serving duties by trying to interpret indirect ways of expressing requests to execute certain actions. In case of uncertainty about the proper interpretation the robot can ask for clarification and adapt its interpretation for future interactions.**

I. INTRODUCTION

Understanding the motives behind what someone says is crucial to successful interaction. This is particularly true in service-oriented human-robot interaction (HRI), where a robot often needs to act on the basis of what the human tells it. The problem is, however, that people don't always literally mean what they say: they say A, but indirectly really mean B – and B is what you ought to do. We say that such utterances express *indirect speech acts* (ISAs) [1]. The issue is nicely illustrated in Asimov's classical story "Little Lost Robot" (1947), where a robot was told "Get lost!" and then subsequently tried to lose itself, rather than just take a step back not to interfere.

In this paper we present an approach towards dealing with ISAs that express *requests*. Examples are questions like "Could you get me a coffee?" (not be taken as a question after the robot's ability), or assertions such as "I am thirsty" (not just about the speaker's physical state, but a request for getting something to drink). Such ISAs occur frequently in human-human interaction [2]. Given the tendency for humans to anthropomorphise machines [3], and their preference for interacting with machines like they do with humans, ISAs thus not only pose a problem for HRI, they are actually desirable for HRI. Interaction with a service robot would be more natural for humans if it were able to handle ISAs, besides direct requests like "Bring me a cup of coffee!".

Our approach to handling request-ISAs starts from a proper analysis of the meaning of an utterance. From this meaning the robot can determine whether it makes sense to interpret an utterance as a request in the given situation [2]. Should the robot be in doubt, the approach enables it to clarify with the user whether an action was requested, and adapt its interpretation mechanisms to handle similar utterances as requests on future occasions. Otherwise, the robot establishes the actions necessary for fulfilling the request, and carries them out. We have completely implemented the approach, and incorporated it into a larger architecture for situated interaction between a human and an ActivMedia PeopleBot.

[4] argue that the kind of pragmatic understanding we need for understanding request-ISAs is crucial to socially interactive robotics. The key characteristics of this approach, namely *situation-dependent flexibility* and *adaptivity*, thus provide an important step towards the ability of the robot to understand what people really mean. The characteristics also set our approach apart from other efforts.

There are two general views on understanding ISAs, namely the *inferential* and the *idiomatic* perspective. The inferential view uses a theory about an agent's beliefs, desires and intentions to infer the intentions behind what she does [5], [6]. Understanding ISAs thus becomes a plan inference problem. The drawback of this is that it can be computationally very expensive. As far as we know, these sophisticated models have never been implemented in a running dialogue system. *Idiomatic* approaches exploit the fact that many ISAs acts are so conventionalized that they are idiomatic, with a fairly fixed form. Most approaches therefore restrict themselves to these fixed-form ISAs, translating them directly into the (prototypical) underlying intention, e.g. [7]. However, such a translation is situation independent making it difficult to handle ambiguous cases or adapt to specific situations.

[8] discuss a dialogue system in which both approaches are implemented, following [9]. The system interprets speech acts relative to the current problem solving context. However this notion of situatedness does not suffice for the situatedness relevant in HRI. We have to deal with situations that not only comprise the current beliefs and intentions of the robot, but also its ability to act in a physical and social environment. Complex dialogue systems in HRI systems like [10], [11], [12] take this broader notion of situatedness into account and would be capable of handling ISAs but so far have not addressed the problem. Simpler HRI dialogue systems based on finite-state approaches lack the flexibility and adaptivity required for situation- and user-dependent interpretation of ISAs.

Contributions. We present a novel approach for interpretating basic forms of request-ISAs in HRI. The novelty is the combination of situation-dependent flexibility, and on-line adaptivity: Interpretation considers the current situation whether to regard an utterance as a request, and the robot can learn on-line how to adapt its interpretation strategies. The approach has been fully implemented in an HRI architecture.

Overview. We present our approach in §II, and its implementation and integration into an architecture for HRI in §III. In §IV we illustrate the approach on running examples, discussing the advantages and shortcomings of the approach. We relate this discussion to other approaches in §V. The paper closes with conclusions and future research.

II. APPROACH

This section describes how we identify request-ISAs. We specify how the interpretation of an utterance as a potential request is based on its linguistic expressed meaning of an utterance, its characterization as a particular type of question or assertion, and the possibility to address the request in the current context.

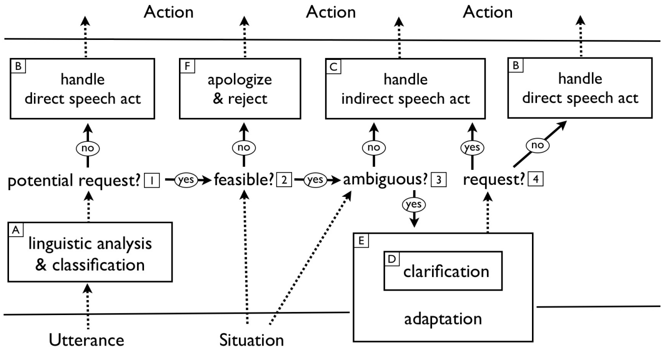

Fig. 1. flow diagram for decision process

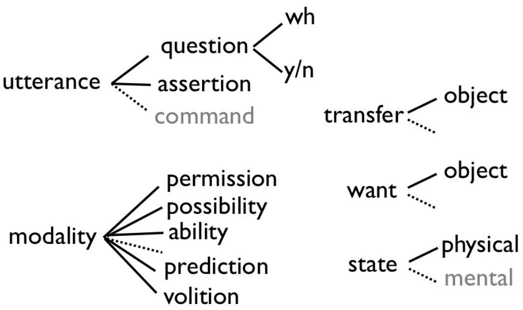

Fig. 2. features for determining the type of an utterance

(1) Questions
  a. Can you bring me a coffee?
  b. Could you bring me a cup of tea?
  c. Will you bring me a coffee?
  d. Would you bring me a coffee?
  e. Can you bring me a coffee, please?

(2) Assertions
  a. I would like to have a cup of tea.
  b. I need tea.
  c. I am thirsty.

The above examples illustrate prototypical request-ISAs we consider. All these utterances could be meant as requests to get something to drink. Whether the robot will actually understand them as requests is determined in an interpretation process which takes into account the linguistic form and the situation in which the utterance was made. Figure 1 illustrates the steps in the interpretation process. First, the utterance will be analysed for its linguistic properties (A), This information allows us to decide if we deal with a potential request (step 1). If not, we treat the utterance as a direct speech act, (either a direct request or a direct act of some other sort) (B); else, we have to check if it is feasible given the situational context (step 2). If the robot would not be able to act according to the request, it will apologize and reject (F). Else, we have to decide if the utterance is ambiguous (step 3). That is we can either handle its indirect meaning right away (C), or we have to clarify the intended meaning before proceeding (D). The decisions in step 2 and 3 are determined by the type of the utterance and the situational context. The situation basically comprehends the robot's ability, readiness and willingness to fulfill the request. (For instance, if the robot had established its readiness to serve the human by uttering "How can I help you?", human's utterances are more likely to contain requests, than in a situation where the robot is obviously busy with the execution of some other actions.) To clarify ambiguous acts the robot asks the human after the intended meaning. Depending on the answer, the robot interprets the utterance either as a direct (B) or indirect (C) speech act. By learning the intended meaning the utterance becomes unambiguous and the robot can adapt its future communication behavior so it need not clarify again but can adopt the learned interpretation.

Now we look a bit closer at the steps in the decision process. Figure 2 shows a subset of types that we are using for the classification of the utterance's meaning. Based on the mood of the utterance we classify it as a *question*, an *assertion* or a *command*. We also determine the modality, and a content type (losely related to SUMO [13]). E.g. (1c) is a *question* with the underspecified modality {*prediction*, *volition*} and content type *transfer.object*. For our decision in step 1 we rely on this analysis, handling *command* utterances as direct speech acts (B), while passing *questions* with modality {*prediction, volition*} and content type *transfer.object* to the next steps.

Decision steps 2 and 3 take the situational context into account. This comprises beliefs, intentions, and abilities of the robot. To fulfill the request "Go to the kitchen!" the robot has to *know* the location of the kitchen and how to get there. The *situation* refers to these requirements, and the current intentions of the robot. The robot can be motivated to serve or it can be occupied with other actions preventing it from fulfilling a new request. We capture this aspect of the situation using a variable that refers to the *mode* of the robot. We distinguish three modes that the robot can be in: The *servant mode* that arises after the robot explicitly offered its services, the *non-servant mode* for when the robot is not ready to serve, and the *default mode* for all other cases. In step 2 we check if the action is feasible given the robot's skills and knowledge and its readiness to act. Given the general feasibility of the action, we check in step 3 if it is ambiguous depending on the robot being in servant or default mode .

If the robot is occupied with other actions that do not allow it to fulfill a request (*non-servant mode*), or if its knowledge or skills are not sufficient, it will not even try to decide if an utterance is to be interpreted as an indirect request. How the robot will react depends on the type of utterance it needs to reply to. For example, to questions as illustrated in (2) it will respond with an apology. Depending on the reason for refusal it will either add an explanation of what it is doing at the moment or a justification that points to missing knowledge or skills. ("Sorry, I do not know where the kitchen is.") This is an appropriate reaction to requests as well as literal questions [2]. Assertions (2) will be answered with a simple acknowledgement ("I see."). This would not be an acceptable answer to a request, but as the robot is not ready or willing or able to fulfill a request anyway, this reaction serves as a good reflection of its state of mind.

The *servant mode* promotes interpreting utterances as requests: All the utterances in (1) and (2) will be taken as request-ISAs and are handled without the need to clarify.

In *default mode*, the reactions are less biased to request interpretation. The robot interprets the assertions in (2) as ambiguous and clarifies their intended meaning. Questions that contain a *please* (1e), are interpreted as requests, because the modifier *please* inhibits the interpretation as literal question. For the questions (1a)-(1d) we distinguish the present tense and past tense forms. This tense information is part of the linguistic meaning of an utterance, which we obtain from linguistic analysis. We do this because past tense forms are less likely to be used for direct questions than present tense forms [2]. For asking if someone is able or allowed to do X, *can* would be more appropriate to use than *could*. The same holds for direct questions after future plans: for asking if someone is planning to do X, *will* is better than *would*. So as *could* and *would* are associated with a greater degree of tentativeness and politeness than *can* and *will*, we assume them to be more likely used in expressing requests. Thus we will interpret the questions built by the past tense form (1b) and (1d) as indirect requests, whereas we regard the meaning of present tense questions (1a) and (1c) as ambiguous and pass them to the clarification process.

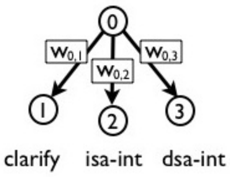

Fig. 3. Adaptation process

For clarification the robot produces a question like: “Do you want me to take this as a request?”. According to the human’s answer the robot adopts the indirect or direct reading and acts accordingly. The robot uses the result of the clarification to adapt interpretation strategy for this pair of utterance type and situation. For each possible pair of utterance type and situation it has three options: it can (1) ask for clarification, (2) adopt the indirect interpretation, or (3) adopt the direct interpretation (cf. Figure 3). Deciding for an option depends on the different weights $w(OUT_{0,i})$ associated with the outgoing edges from 0. More specifically, we take the option /empho with the maximal weight: $o = argmax_{i=1,\ldots,n}(w(OUT_{0,i})$ with $i \in \{1,2,3\}$

Adaptation proceeds by adapting the weights with an instantiation of the following general adaptation function: $f(w(OUT_{0,i}), o, answer) = w'(OUT_{0,i})$, normalizing the sum of all weights to 1. In the adaptation function the weights $w_{0,i}$ only change depending on the answer to a clarification request. For the purpose of this paper we adopt a simple step function that leads to a stable system in only one learning step, given the initialization of $w(OUT_{0,1}) = 1$ and $w(OUT_{0,2}) = w(OUT_{0,3}) = 0$.

$$f(w(OUT_{0,i}), 1, \text{“}yes'') = \left\{ \begin{array}{lcl} 1 & : & i = 2 \\ 0 & : & i \neq 2 \end{array} \right.$$

$$f(w(OUT_{0,i}), 1, \text{“}no'') = \left\{ \begin{array}{lcl} 1 & : & i = 3 \\ 0 & : & i \neq 3 \end{array} \right.$$

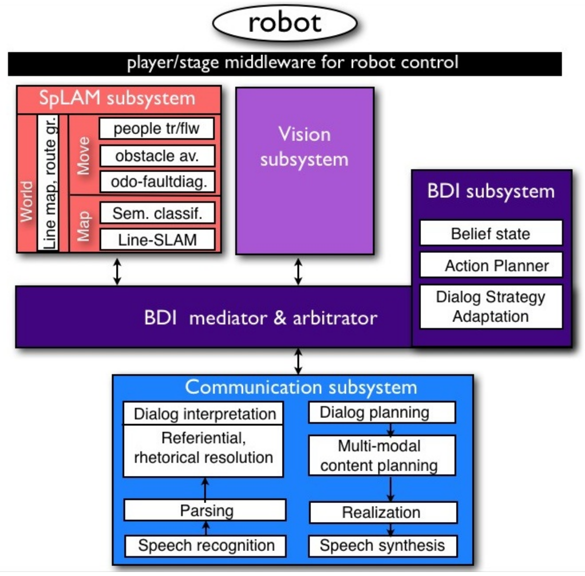

Fig. 4. architecture

To avoid overfitting and to ensure flexibility to possibly changing environments, we allow the robot to choose option 1 (i.e. clarification) instead of a maximally weighted option 2 or 3 with a probability of 0.1

We illustrate our approach in §IV, after having discussed its implementation within the larger architecture for HRI.

III. IMPLEMENTATION

We have implemented the approach of §II in a distributed architecture which integrates different sensorimotoric and cognitive modalities. The architecture enables a robot to move about in an indoor environment, and have a situated dialogue with a human about various visual and spatial aspects of a situation. We have deployed this system in scenarios for human-augmented mapping and simple visual object manipulation, using an ActivMedia PeopleBot.

Figure 4 shows the relevant aspects of the architecture. We have subsystems for communication, spatial localization & mapping, visual processing. We use the BDI-subsystem as a mediator between the other subsystems. By this we mean that beliefs provide a common ground between different modalities, rather than being a layer on top of the different modalities. Beliefs provide a means for cross-modal information fusion, in its minimal form by co-indexing references to information in individual modalities [14]. It is at this mediating level that we deal processing speech acts, and that we handle requests to clarify issues that have arisen in a particular modality, through communication or through another modality.

The communication subsystem consists of several components for the analysis and production of natural language. It has been implemented as a distributed architecture using the Open Agent Architecture [15], following the idea of agent-based models for multi-modal dialogue systems [16]. On the analysis side, we use the Sphinx4 speech recognition engine[1] with a domain-specific speech grammar. The string-

---

[1] `http://cmusphinx.sourceforge.net/sphinx4/`

based output of Sphinx4 is then parsed with OpenCCG[2]. OpenCCG uses a combinatory categorial grammar [17] to yield a representation of the linguistic meaning for the recognized string/utterance. We represent linguistic meaning as an ontologically rich, relational structure in a description logic-like formalism [18], e.g. see (4) below. In dialogue analysis we relate the linguistic meaning of an utterance to the current dialogue context, in terms of how it rhetorically and referentially relates to preceding utterances. This yields an updated model of the situated dialogue context [19], [11].

To produce flexible, contextually appropriate interaction we use several levels of dialogue planning. Based on a need to communicate, arising from the current dialogue flow or from another modality, the dialogue planner establishes a communicative goal. We then plan the content to express this communicative goal, possibly in a multi-modal way using non-verbal (pose, head moves) and verbal means. In these planning steps, we can inquire the models of the situated context (e.g. dialogue context, visually scene) to ensure that the content we plan is contextually appropriate. We realize verbal content using the OpenCCG realizer, which generates a string for the utterance, and then synthesize this string using text-to-speech[3].

The spatial localization & mapping (SpLAM) subsystem first of all maintains information about the spatial organization of the environment. The robot uses this information as the basis for planning its actions in the environment. The subsystem is based on metric SLAM, with which the robot can create a feature-based map of the environment [20]. While the robot moves around the environment, we use a technique as in [21] to also build a routegraph on top of the feature-based map. This routegraph captures the free space in the environment, and the edges in the graph represent paths for moving from one place to another. The routegraph provides the basis for a topological graph in which each area in the environment corresponds to a single node, with edges representing which areas are connected. When a human guides the robot around a new environment, and tells the robot more about the areas it is in, the robot stores these area descriptions in this topological graph. Other functionality the SpLAM subsystem provides includes people following, based on dynamic object tracking, obstacle avoidance, and odometric fault diagnosis.

The BDI-subsystem processes utterances with their linguistic analysis coming from the communication subsystem. As we described in §II it decides based on contextual information whether to treat the utterance as a direct speech act and trigger the appropriate subsystems or if an indirect reading is possible. For clarification it triggers the communication subsystem to start a clarification dialogue. The BDI-subsystem then handles the received results in its module for dialogue strategy adaptation and triggers an appropriate reaction.

The next section illustrates those processes in more detail.

IV. EXAMPLES

In this section we illustrate our approach on several example runs. The examples illustrate how the robot can use access to action & perception-related modalities to achieve situation-dependent flexibility in interpreting speech acts, and how it can adapt its interpretation strategies through clarification dialogues with the human user. The actions of the robot that follow from some speech acts are planned by the action planner module in the BDI subsystem and enabled by the subsystems described in §III.

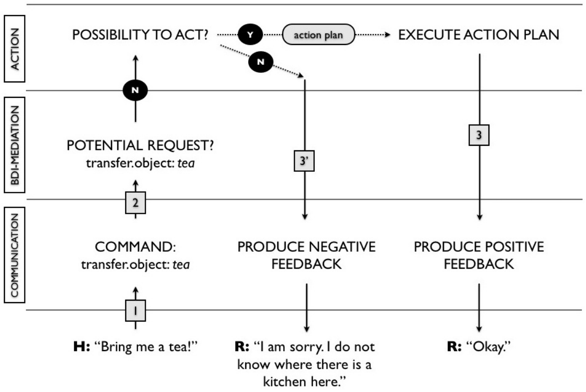

Fig. 5. Process flow diagram for "Bring me a tea!"

We start with an example of a direct speech act (a command), to illustrate some basic mechanisms of the architecture:

(3) "Bring me a tea!"

Figure 5 illustrates how the command in (3) is processed. When the communication subsystem receives the speech signal for "Bring me a tea!" (step 1), it parses the string to obtain a representation of the linguistic meaning it expresses:

(4)
$$\begin{aligned} &@b1 : action(bring \wedge \langle Mood\rangle imp \wedge \\ &\quad \langle Actor\rangle(r1 : hearer \wedge robot) \wedge \\ &\quad \langle Patient\rangle(t1 : thing \wedge tea \wedge \\ &\quad \langle Recipient\rangle(i1 : person \wedge I)) \end{aligned}$$

(4) shows that the utterance expresses a **bring**-action, using *imp(erative)* mood. The robot is the one to perform the action (*Actor*), getting the tea (*Patient*) for "me" (*Recipient*). We determine the utterance to be a *command to transfer an object* on the basis of the action and the mood. When the BDI mediator receives this information (step 2), we can immediately interpret the utterance as a direct speech act. We now need to check whether we can build an action plan. If we can, then we execute the plan, and tell the BDI mediator that we were successful. The BDI mediator in turn triggers the communication subsystem to produce a positive feedback to indicate that we have understood the command and are carrying it out (step 3). If we cannot form a plan in the current situation, then a negative indication is sent back to the BDI mediator, possibly with a reason why we cannot form a plan. In this case we produce a negative feedback (step 3').

Figure 6 shows how we process the request-ISA in (5):

(5) "Can you bring me a tea?"

The utterance in (5) also expresses a **bring**-action, but now in *interrogative* mood with a modal "can". Based on the types in Figure 2 we analyze the uttterance (5) as a *question* with the

---

[2] `http://openccg.sf.net`

[3] `http://mary.dfki.de`

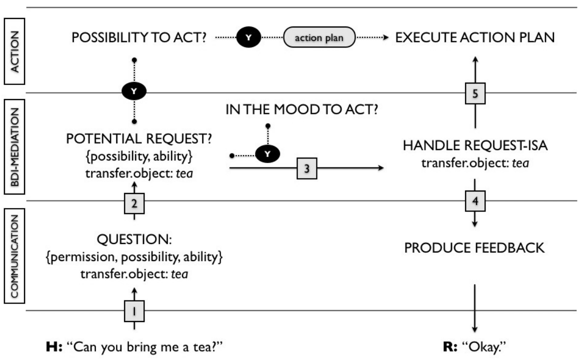

Fig. 6. Process flow diagram for "Can you bring me a tea?"

underspecified modality {*permission,possibility,ability*} and the semantic type *transfer.object* (step 2). This type qualifies it as a potential request.

We first check whether it is sensible to interpret the question as a possible request, based on the ability to execute the requested action, and the robot's mode. This illustrates the situation-dependent interpretation of requests. positively, and that the request is unambiguous, If the robot is able to form an action plan and it is in *servant mode* the request is deemed unambiguous and the BDI mediator handles it (step 3) by triggering the execution of the formed action plan (step 5), and the communication of a positive feedback (step 6). If we cannot form an action plan in the current situation, we proceed as in step 3' in Figure 5. If the robot is in *default mode* (cf. §II), the request is ambiguous and needs further clarification. (See also the following example.)

Finally, we illustrate how we deal not only with situation-dependent interpretation, but also how we adapt interpretation strategies should an ambiguity arise in how we should interpret a (potential) request-ISA. Consider for example (6):

(6) "I need a cup of tea."

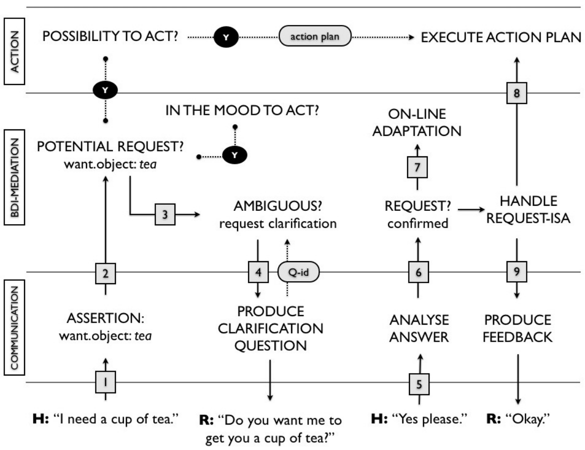

Fig. 7. Process flow diagram for "I need a tea."

Figure 7 gives the process flow for (6). The communication subsystem analyzes the utterance as an *assertion* of type *want.object*, on the basis of its *indicative* mood and **need**-predicate. This presents a potential request to the BDI mediator (step 2). Now assume that the robot knows how to fulfill this need (i.e. it can form an action plan), but that it is in *default mode* (cf. §II). Thus, it could in principle handle the request, but it is not sure whether to do so (as no explicit serving mode has been triggered earlier). The robot considers the request ambiguous (step 3), and in need of further clarification. The BDI mediator sends a request to the communication subsystem to raise the issue with the human user. The communication subsystem produces a clarification question: "Do you want me to get you a cup of tea?" and stores this question in its model of the dialogue context, while returns the identifier of the question to BDI mediator. When the human answers the question (step 5), the rhetorical analysis in the communication subsystem can tie the answer to the previously posed clarification question.

The communication subsystem informs the BDI mediator of the answer, and the identifier of the question it was an answer to (step 6). Through the question-id the BDI mediator can use the answer to resolve the ambiguity issue of the potential request-ISA: The human's positive answer confirms the utterance as a request.[4] We use the confirmation for two purposes: to handle the request, and to on-line adapt the interpretation strategies for utterances like (6). To handle the request, we execute the formed action plan (step 8) and produce a positive feedback (step 9).

Referring to the adaptation function described in §II, the weights for the different interpretation strategies have changed now. At initial state *clarify* was the option with the highest weight, the positive answer of the human changed the weights such that the interpretion as an indirect speech act is now maximally weighted. The next time that the robot has to interpret an utterance of the same type in the same situation it will choose to adopt the indirect meaning as a request with a probability of 0.9 and to pose a clarification question again with a probability of 0.1.

V. DISCUSSION

Our approach to handling ISAs is distinct from others in that it provides an adaptative component for ambiguous acts and that it takes into account the influence of the situation in which the speech acts are employed.

In general there are two approaches for dealing with the recognition of indirect speech acts, the *inferential* and the *idiomatic*. The former uses a complex theory about the agent's beliefs, desires and intentions, and uses those to infer why the agent performs certain acts ([5],[6]). These logical approaches analyse speech acts similar to other actions assigning preconditions and effects. From this perspective producing a speech act is a planning problem and recognizing a speech act is a plan recognition (or plan inference) problem. The drawback of these approaches is that they can be very expensive. To our knowledge those sophisticated models have never been implemented in a running dialogue system.

The *idiomatic* approaches exploit the fact that many indirect speech acts are so conventionalized that they have become

---

[4] Alternatively, if the human denies the indirect interpretation, the robot interprets the question literally and reacts verbally, e.g. by answering: "I could if you want me too."

idioms. As a result, it is hardly possible to interpret them as a direct speech act. These approaches are restricted to idiomatic cases. [7] gives an example where idiomatic indirect speech acts are translated into their direct meaning. However translation is restricted to indirect questions like “I do not know X” or “I would like to know X” being translated to “X?”. The translation is situation independent making it difficult to handle ambiguous cases or adapt to specific situations.

[9] combines both approaches, [8] implement it in the practical interactive planning system TRIPS. The system interprets speech acts with regard to the current problem solving context. However this notion of situatedness is different from the situatedness relevant in an autonomous robotic scenario. We have to deal with situations that comprise the current beliefs and intentions and abilities of the robot in its current physical and social environment.

Dialogue systems in HRI systems like [10], [11], [12] seem to have the complexity to handle ISAs but to our knowledge they have not addressed the problem. management on an autonomous companion robot similiar to ours that uses use finite state machine technology.

Our approach to interpret utterances concentrates on the linguistic analysis and the situational context. We do not yet have an explicit model for intentions and collaborative aspects [10], which we need to integrate the problem solving context [8] into our notion of situation. Furthermore, the way we handle ISAs can be improved in other ways. So far we only deal with requests-ISAs, and could thus extend to handle other types of ISAs that are relevant in the service robotic scenarios. Also, although the current system can adapt its interpretation of certain types of utterances in certain sitations it still lacks a general adaptability regarding all possible interpretations. Utterances that are not classified as ambiguous will always be interpreted according to the supposed model. If this interpretation clashes with the intention of the user, it is impossible to resolve this misunderstanding and to teach the system the more appropriate interpretation. To extend the adaptational mechanism to those cases, the dialogue management needs to be refined such that it enables the resolution of those misunderstandings. Finally, if we could distinguish between humans (e.g. through face recognition), we could make the adaptation user-specific thus learning user-specific preferences.

VI. CONCLUSION

We described an approach to interpret request-ISAs in human robot interaction. The robot interprets utterances based on their linguistic meaning and classification in the context of the current situation. If an utterance is initially ambiguous the robot asks for clarification and adapts its interpretation strategy. The key features of our approach are *situation-dependent flexibility* and *adaptivity*. We implemented and tested the approach on a real robot that can move autonomously in an indoor environment. It can fulfill direct or indirect requests for bringing objects by obtaining objects from locations where it knows such objects can be found.

ACKNOWLEDGMENTS

The authors would like to thank ...

REFERENCES

[1] J. R. Searle, *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, United Kingdom: Cambridge University Press, 1969.

[2] S. C. Levinson, *Pragmatics*. Cambridge University Press, 1983.

[3] C. I. Nass and Y. Moon, “Machines and mindlessness,” *Journal of Social Issues*, vol. 56, no. 1.

[4] T. Fong, I. Nourbakhsh, and K. Dautenhahn, “A survey of socially interactive robots,” *Robotics and Autonomous Systems*, vol. 42, no. (3-4), pp. 143–166, 2003.

[5] C. R. Perrault and J. F. Allen, “A plan-based analysis of indirect speech acts,” *American Journal of Computational Linguistics*, vol. 6, no. 3-4, pp. 167–182, 1980.

[6] A. Herzig, D. Longin, and J. Virbel, “Towards an analysis of dialogue act sand indirect speech acts in a bdi framework,” in *Proc. 4th Workshop on the Semantics and Pragmatics of Dialogue (GTALOG 2000)*, M. Poesio and D. Traum, Eds., 2000.

[7] M. Gerlach and M. Sprenger, “Semantic interpretation of pragmatic clues: connectives, modal verbs, and indirect speech acts,” in *Proceedings of the 12th conference on Computational linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 1988, pp. 191–195.

[8] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, “Towards conversational human-computer interaction,” *AI Magazine*, vol. 22, no. 4, pp. 27–37, 2001.

[9] E. Hinkelman and J. Allen, “Two constraints on speech act ambiguity,” 1989.

[10] C. L. Sidner, C. Lee, C. D. Kidd, and N. Lesh, “Explorations in engagement for humans and robots,” in *IEEE RAS/RSJ International Conference on Humanoid Robots*, November 2004.

[11] J. Bos, E. Klein, and T. Oka, “Meaningful conversation with a mobile robot,” in *Proceedings of the Research Note Sessions of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL’03)*, Budapest, Hungary, 2003.

[12] O. Lemon, A. Bracy, A. Gruenstein, and S. Peters, “The witas multi-modal dialogue system i,” in *EUROSPEECH-2001*, pp. 1559–1562.

[13] I. Niles and A. Pease, “Towards a standard upper ontology,” 2001.

[14] I. Gurevych, R. Porzel, E. Slinko, N. Pfleger, J. Alexandersson, and S. Merten, “Less is more: Using a single knowledge representation in dialogue systems,” in *Proceedings of the HLT-NAACL WS on Text Meaning*, Edmonton, Canada, 2003.

[15] A. Cheyer and D. Martin, “The open agent architecture,” *Journal of Autonomous Agents and Multi-Agent Systems*, vol. 4, no. 1, pp. 143–148, March 2001.

[16] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, “An architecture for a generic dialogue shell,” *Journal of Natural Language Engineering*, vol. 6, no. 3, pp. 1–16, 2000.

[17] J. Baldridge and G.-J. M. Kruijff, “Multi-modal combinatory categorial grammar,” in *Proceedings of EACL 2003*, Budapest, Hungary, 2003.

[18] ——, “Coupling CCG and hybrid logic dependency semantics,” in *Proceedings of ACL 2002*, Philadelphia, Pennsylvania, 2002.

[19] N. Asher and A. Lascarides, *Logics Of Conversation*, ser. Studies in Natural Language Processing. Cambridge, United Kingdom: Cambridge University Press, 2003.

[20] J. Folkesson, P. Jensfelt, and H. Christensen, “Vision slam in the measurement subspace,” in *Proceedings of the IEEE International Conference on Robotics and Automation*. Barcelona, Spain: IEEE, 2005.

[21] P. Newman, J. Leonard, J. Tardós, and J. Neira, “Explore and return: Experimental validation of real-time concurrent mapping and localization,” in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA’02)*, Washington D.C., USA, 2002, pp. 1802–1809.