The Language of Machines

Eva Jacobus Brian Duffy

Media Lab Europe (MLE), Sugar House Lane, Bellevue, Dublin 8, Ireland

E-mail: ejacobus@media.mit.edu brd@media.mit.edu

Abstract. *When we start to engage machines in conversation, what language will it be? While natural language is our preferred means of communication, natural language processing is by no means trivial. An alternate approach is to look at augmenting Agent Communication Languages with a view to providing a middle ground between for speech-based interaction between man and machine. This paper presents such a system which has been implemented on Media Lab Europe's social robot "Joe".*

1 Introduction

As machines become pervasive, numerous, and even autonomous fixtures in the home and workplace, our interactions with them will become more sophisticated and inevitable. The machine's social interaction with us and other machines will develop as explicit social interactions in order to manage the increasing number of entities (agents) sharing the environment to accomplish their tasks. This work seeks to investigate the development of social interaction between man and socially capable machines, particularly with regards to a communication language suitable for both machines and people to support communication in the new, shared human-machine environment.

2 Social Interaction with Machines

In recent years, examples of explicit social interaction between machines has been framed in terms of a multi-agent system, where individual agents pursue their own interests or share tasks to accomplish a collective goal [Duf00] [Col01]. In social robotics, communities of multiple mobile agents can decompose a complex task, and negotiate the performance of its constituent steps through the use of a structured language of communication [ROD+99].

An agent in the context of a robotic community is an embodied autonomous unit possessing its own goals, decision-making processes, and physical capabilities. An agent's goals may be individually held or communally shared, and an agent may approach a conflict between individual and community goals with competitive and cooperative strategies [Mue96]. Moreover, an agent's physical abilities may or may not be sufficient to complete a goal without the aid of other members of the community.

The benefit of the community lies in increasing the complexity of achievable goals by widening the range of skills available to the group, without necessitating a corresponding increase in the complexity of the individual members of the community. As in many biological societies, delegation and specialization increase the productivity of the group compared to the productivity of multiple individuals. The issues of how to develop and maintain these social interactions come to the fore.

When a human is added to the social loop, a basic form of human-machine interaction can be defined as the acceptance of the human as an agent by the machine within a multi-agent system. Interaction, particularly goal-oriented interaction, in a community of agents requires both a communication protocol between agents, and the availability of a profile of each agent's abilities to the community, so that negotiation of actions may take place. A human can enter into a community of machines, provided that he shares a communication protocol with the machines and that a profile of the abilities of a generic or specific human agent can be shared with the community.

3 Robot Communication and Language

Existing communication methodologies in robotics can often be classified as either using explicit communication to pass state information between component robots, or, in the context of a reactive robot's control structure, building rules for such interactions. Explicit communication can easily fail when confronted with no predefined communication protocol or when external dissimilar agents are introduced into the system. Reactive approaches on the other hand, while robust and flexible in their limited domains, are too simplistic for use in domains that require more complex reasoning. It can be difficult to extend reactive systems to complex domains, and they frequently suffer from dysfunctional emergent behaviour when their rule-bases become complicated [Bro86].

Language not only serves to acquire knowledge about behavioural characteristics of others, but also to find out their internal states (i.e. their feelings, attitudes, etc). In order to build up a basis for interaction and co-operation, individuals have to communicate and merge their conceptions of the world (i.e. world models), with the degrees of abstraction being socially grounded and continuously updated. In addition to being perceived as a means of communicating ideas, knowledge and experience, language is also a tool that alters the nature of the decision making process [MP43].

Communication in multi-robot systems ranges from none at all to interaction through high-level agent communication languages (ACLs). Among the approaches for robot communication are; none [KZ93], implicit communication through the environment [DJM+95], simple semaphores or signals [BA94], simple message passing [BA94], and Agent Communication Languages [LF94] [FIP97] [Roo00]. The level of communication complexity has generally been representative of the complexity of the robot systems in terms of both its interactive and computational capabilities.

4 Agent Communication Languages (ACL)

Complex inter-agent task-based interactions require a language with sufficient expressive power to represent and express the concepts of beliefs, intentions, requests for information and services, replies to such requests, etc. Agent Communication Languages have developed with sufficient complexity to facilitate this where an ACL is “a *language with a precisely defined syntax, semantics and pragmatics that is the basis of communication between …agents.” FIPA [FIP97]*

Agent Communication Languages are high-level languages used by agents for the exchange of information and requesting services. They generally have a syntax powerful enough to generate a wide range of communicative actions. There are a number of features, which an ACL should possess, and criteria by which they may be judged. Mayfield *et al.* [MLF95] present one such list of *desiderata* for agent communication languages. In brief:

***Form**:* An ACL should ideally be syntactically simple, concise, yet extensible.

Content:A distinction should be made between the language, which expresses communicative acts, and the content language.

***Semantics**:* The semantics of a communication language should be grounded in theory and unambiguous.

***Implementation**:* The implementation should be efficient, provide an intuitive interface that hides the message transport details, and be amenable to partial implementation.

***Networking**:*The language should be suitable for implementation on top of modern networking technologies.

***Environment**:*An ACL should be designed for a heterogeneous and highly dynamic environment.

***Reliability**:* The language must be able to provide secure and reliable message transport.

4.1.1 Speech Act Theory

Speech Act Theory, widely used in ACL's is founded on the notion that one not only make statements, but also performs actions [Aus69]. For example, when one requests something, one does not just report on a request, but one actually *effects* the request. That is, the speech act not only reports but also involves a change of state. For most computing purposes, speech acts are classified into assertives (informing), directives (requesting), commissives (promising), permissives, prohibitives, declaratives, and expressives (i.e. for emotions). Existing theories treat utterances as either true or false, i.e. as *constatives*. Austin specified a new class of utterances, *performatives*, which was not just for making statements, but rather to actively do things. Performatives are considered speech *acts*, like any physical act such as pushing; they cannot be true or false but they can fail. Performative verbs are verbs that may be used to perform actions, e.g., request, promise, assert. It is with this theory in mind that the ACL Teanga was developed.

4.1.2 Teanga

Teanga [Roo00] [ROD+99] is based upon Speech Act Theory [Aus69]. It consists of 4 basic categories of *communicative acts* (messages): informatives, directives, commissives and declarations. This categorisation was developed from a classification of performatives proposed by [Sea76]. Within each category are more specialised types of communicative acts, e.g. `drop_commitment`. The language is designed as a carrier for an application-dependent content language. Some constraints are, however, placed on prospective content languages. In the context of the social interaction between agents the content language must be able to allow, amongst other things, the representation of actions (including speech acts) and their status e.g. done, doing; and an agent's mental states, e.g. beliefs and commitments.

The difference between Teanga and the established languages such as KQML [LF94] or FIPA's ACL [FIP97] is the requirement of a language that is compositional. This allows the support of nested speech acts (and speech acts contained inside composite actions, e.g. plans). Examples of the communication structure follow. The first shows the generic form of a speech act represented in EBNF. The other examples are fully formed communicative acts. The generic structure for a communicative act is as follows:

```
<speech act>::=SPEECH ACT (<sender><recipients>) (<structure>)
```

The language *Teanga* allows the development of explicit high-level conversation between robots in a social community, an attribute of inter-robot sociality that has been developed or implemented using the Social Robot Architecture [Duf00].

5 An ACL for a human-inclusive Multi-Agent System

When a human enters the social loop, the form of the social interaction becomes more complex. For humans and machines to function in the same multi-agent system, they must possess a mental model of the abilities of the other types of agent in their environment and a protocol with which to communicate. A machine possessing anthropomorphic qualities is more readily stereotyped as socially capable by the human, based on human preference for social interaction and tendencies to project human-like internal motivations where human-like external characteristics are present [Duf03]. However, the issue of a common language of interaction is less easily approached.

5.1.1 Need for an extended ACL

In defining a human-machine interaction as a type of multi-agent system, the need to resolve differences in linguistic sophistication immediately emerges as an obstacle to the communication protocol necessary for interaction between agents. The current ability of machines to understand or produce human speech is not sufficient for communication between man and machine in man's language. Producing human speech demands repetitive and simplistic mimicry of basic grammatical forms, which often results in a grammatical (and sometimes logical) tangle. Understanding human speech is similarly dependent on grammatically limited phrases, where the human bears the additional burden of using precise vocabulary and grammatical form in a linguistic medium where it seems unnatural.

Fortunately, framing the interaction as a multi-agent system also suggests the solution to the communication problem. An Agent Communication Language that mediates the differences in the abilities of the agents in the system increases the robustness of the machine's participation through its defined concise syntax. Work being done at Carnegie Mellon University echoes the need for a communication language in which "the user and the machine meet halfway" and in which, "[g]iven the limitations of speech recognition and language processing", the interface conveys "to the user the fact that their conversational partner is simply a tool" [SRZ+01].

5.1.2 Needs of an extended ACL

Machine agents can require a simple and precise grammar to avoid mistakes in parsing incoming phrases and in creating grammatical responses. By reducing the complexity in the language, one reduces both the demands on the less linguistically sophisticated agent (the robot) and the expectations of the more linguistically sophisticated agent (the human). However, incompleteness in utterance structure must be expected and explicitly handled by the machine, as part of a model of human speech preferences, such as brevity and the tendency to avoid volunteering information deemed excessive [Gri75], and the grammar and vocabulary must be sophisticated enough to bear a resemblance to the natural language of the human agent so as to be accessible to the human. This need for similarity to natural language is tempered by the fact that much of human language is conversational, and what is required here is not conversation, but a task-based command set with data attributes. The needs of an extended ACL, therefore, include natural language vocabulary, precise, unambiguous grammar, application-specific modifying information, and the ability to handle incomplete commands.

The most important feature of a designed grammar, both in terms of human learning and machine parsing, is a lack of syntactic ambiguity. Our syntax is derived from that of Teanga [Roo00], consisting of a wrapper that contains the speech act being communicated, the intended recipient, and the enclosed application-specific information. The enclosed information, being variable in length and type, is the most likely to be forgotten or given out of order by the human participant. Here, the need for unambiguous syntax is served by verbal tags denoting the grammatical role of the following information. The handling of incomplete communications is aided by the verbal tags, which allow the machine to parse partial information correctly and request the missing information to complete the communication. The vocabulary used is abstracted from Teanga, with the goal of achieving more concise phrases, based on Grice's maxim that humans tend to brevity in speech. In particular, where Teanga's speech act subtypes consist of multiple words, we allow the substitution of one-word alternatives, such as "respond" instead of "request answer".

5.1.3 Implementation of an extended ACL

The extended ACL consists of three parts of speech: the utterance type, the target agent name, and the content. The action type can be optionally modified by the addition of properties specific to each action type. The utterance types, based on Teanga's speech acts, consist of "query", "inform", "commit", and "declare", and their subtypes, with shorter versions of commonly used subtypes optional. The target agent name is mapped to a representation of the agent's characteristics, based on whether it is a human or digital agent. This allows the parameter provided as "agent name" to be either the name of a specific agent, or a generalization such as "any human agent". These combined form the wrapper for the content, which can consist of an action type, a property, or an action modified by additional properties. Properties describe the state of the variable or object they are mapped to, and can be required parts of action types or expressed independently as an assertion of belief.

Figure 1: Media Lab Europe's social robot "Joe"

The action types are, as in Teanga, high-level actions, leaving implementation to the individual agent and are bundled with application-specific information, or modifying properties. The modifying properties for action types consist of any piece of information that affect how (or when, or where, if applicable) an action is to be performed. They can be specific to the actions with which they are associated, but are generally of the same types for similar actions, like the filename property associated with opening and closing files, and can be optional or mandatory. For instance, in the case of opening a file, the filename is considered a mandatory property, as it must be supplied for the action to be performed. In the case of listing the files in a directory, the option to list files via text-to-speech rather than via text is considered an optional property.

Examples of speech acts:

```
"Request target Joe action list files property aloud."
"Request target Joe action open file."
"Reply target human-user property filename null."
```

The ACL is currently implemented in C#, and running on three platforms at Media Lab Europe, including two anthropomorphic robots. The speech recognition component uses the Microsoft Speech SDK 5.1, which makes use of a combination of static and dynamic sets of recognition rules. The text-to-speech component uses the Festival Speech Synthesis system, developed by the Centre for Speech Technology Research at the University of Edinburgh. Given that interaction is goal-directed, success in use is viewed as the completion of a task that either stands alone or is a step in a larger goal. Preliminary usage of a limited vocabulary shows a highly accurate recognition and action performance by the robot, even when commands are initially given with incomplete information such as failure to provide filenames. The ability to omit modifying information appears to shorten the learning curve for human users first using the system, as they can concentrate on the accuracy of shorter commands without needing to devote much attention to the modifying content, which they will be prompted for when the robot requests the additional information. Human familiarity with the ACL syntax is the factor with the greatest effect on the percentage of successful interactions, which emphasizes the need for the ACL to support the clarifications of partial or malformed commands.

6 Conclusions

As agents evolve from pure functional machines such as robots in production assembly operations to more socially capable entities in our homes, this necessitates a robust social interaction methodology. Intuitively, aiming to achieve a mutual common ground that facilitates social interaction between artificial systems and humans can reduce ambiguity and provide for more robust communication. The implementation of an Agent Communication Language through speech synthesis and recognition with little modification has proved successful in providing ease of access to a robot's capability set for a human user.

7 References

[Aus69] Austin, J.L., How To Do Things With Words, *Oxford University Press, 1969.*

[BA94] Balch, T., Arkin, R.C. "Communication in reactive multiagent robotic systems", *Autonomous Agents, 1, P1-25, 1994.*

[Bro86] Brooks, R.A., "A Robust Layered Control System for a Mobile Robot", *IEEE Jour. Rob. and Autom., 2(1) 1986*

[Col01] Collier, R.W., Agent Factory: A Framework for the Engineering of Agent-Oriented Applications, *Ph.D. Thesis, University College Dublin, Ireland, 2001.*

[CL95] Cohen, P.R. Levesque, H.J. "Communicative Actions for Artificial Agents". *Proc. Int. Conference on Multi-Agent Systems, AAAI Press, San Francisco, June, 1995.*

[DJM+95] Dudek, G., Jenkins, Milios, E., M., Wilkes, D. "Experiments in Sensing and Communication for Robot Convoy Navigation" *in Proceedings IEEE International Conference on Intelligent Robots and Systems (IROS), 1995.*

[Duf00] Duffy. B.R. The Social Robot. PhD Thesis, Department of Computer Science, University College Dublin (UCD), Dublin, Ireland, 2000.

[Duf03] Duffy, B.R., "Anthropomorphism and The Social Robot", *Special Issue on Socially Interactive Robots, Robotics and Autonomous Systems 42 (3-4), 31 March 2003*

[FIP97] FIPA 97 Specification: Agent Communication Language. Foundation for Intelligent Physical Agents, 1997, *http://drogo.cselt.stet.it /ufv/leonardo/fipa/spec/*

[Gri75] Grice, H. "Logic and Conversation," Syntax and Semantics, Vol 3: Speech Acts. Academic Press, New York, 1975. 41-58.

[KZ93] Kube, C.R., Zhang, H. "Collective Robotics: From Social Insects to Robots" *Adaptive Behavior, 2(2):189-219, 1993.*

[LF94] Labrou, Y., Finin, T., "A Semantics Approach for KQML- A General Purpose Communication Language for Software Agents". *Third International Conference on Information and Knowledge Management (CIKM'94), November 1994*

[MLF95] Mayfield, J., Labrou, Y., Finin, T. "Desiderata for Agent Communication Languages", *AAAI Spring Symposium on Information Gathering, 1995.*

[MP43] McCulloch, W., Pitts, W. "A Logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Geophysics 5, P115-133, 1943*

[Mue96] Mueller, H. Juergen. "Negotiation Principles", in Foundations of Distributed Artificial Intelligence, *ed. O'Hare, G.M.P., and Jennings, N. R. 1996.*

[ROD+99] Rooney, C.F.B., O'Donoghue, R.P.S., Duffy, B.R., O'Hare, G.M.P.,Collier, R.W. "The Social Robot Architecture: Towards Sociality in a Real World Domain" *in Proc. Towards Intelligent Mobile Robots 99 (Bristol). Tech. Report Series, Dept. Computer Science, Manchester University, Rep. No. UMCS-99-3-1. 1999.*

[Roo00] Rooney, C., "A formal semantics for Teanga", Technical Report No. UCD-PRISM-00-05, Department of Computer Science, University College Dublin, 2000

[Sea76] J. R. Searle "The Classification of Illocutionary Acts". *Lang. in Society, 5, p1-24, 1976*

[SRZ+01] Shriver, S., Rosenfeld, R., Zhu, X., Toth, A., Rudnicky, A., Flueckiger, M. "Universalizing Speech: Notes from the USI Project" in Proc. Eurospeech 2001.