

# Social Agents for Communicative Tasks\*

From: AAAI Technical Report WS-96-02. Compilation copyright © 1996, AAAI (www.aaai.org). All rights reserved.

David Pautler

Institute for the Learning Sciences

1890 Maple Avenue

Evanston, IL 60201

pautler@ils.nwu.edu

## Abstract

Agents which interact with humans on behalf of humans need a model of human social interaction in order to act appropriately. This paper is a brief description of such a model used by an agent to generate e-mail messages.

## 1. The problem

There is a great deal of research in AI on multi-agent systems [Tambe et al. 1995; Etzioni, Lesh, & Segal 1993]. But there seems to be a growing need for agents which can interact not only with other agents but with humans as well. For example, the rapidly growing popularity of e-mail has already made it difficult for some people to respond appropriately to all the messages they receive. Reading a user's e-mail and generating appropriate responses is a task seemingly well-suited to intelligent, independent agents. Yet, if an agent is to act as an intermediary between people, the agent must have knowledge of how humans interact with each other in order for the agent's messages to be socially appropriate and productive. Current research in multi-agent systems and in discourse processing does not offer a model of human social interaction. This paper presents a brief overview of such a model and of an agent (a "proxy") which uses the model to generate socially appropriate e-mail messages.

In order to clarify what I mean by "socially appropriate" messages and show how a proxy could fail to meet that criterion, I will start with an example of a message which *is* socially appropriate. The context for the example is that one has been invited to speak before an academic department about one's research, but the scheduled time of the speech conflicts with a previous commitment, so one must decline. The note in Figure 1 is an appropriate response to such an invitation because it shows a fitting level of

gratitude for the offer, in addition to communicating the basic message that the writer cannot accept the offer.

Without the knowledge that a human should show gratitude for offers he or she receives, an agent acting on behalf of such a person might respond merely, "I decline your offer." Upon receiving such a curt response, an inviter would probably feel insulted that the invitee did not value the invitation. One cannot entirely avoid the negative impact of declining, but there are a number of things one can say to lessen the negative impact, such as thanks, apologies, and excuses. Furthermore, an inviter expects such niceties to be observed, so even if the overtures do not change his mind, he is likely to believe that the invitee was acting in good faith. Thus, an agent which interacts with humans on behalf of a human must have *some* knowledge of appropriate social communication (even if that knowledge is implicit in a procedure or data structure, e.g. a decision tree of response-letter templates indexed by the category of message received).

Dear Prof. WHITNEY,

Thank you for your invitation. Unfortunately, I will not be able to give a talk at THE U OF M COMPUTER SCIENCE DEPARTMENT on APRIL 14, 1996. I regret that I must decline. I have a previous commitment.

You may want to invite DAN VLASIK in my place. He is well-acquainted with the work we do here at McCORMICK SYSTEMS. If you would like to pursue this option, please contact him directly at (847) 467-1973.

## Figure 1

A socially appropriate letter for declining an invitation

\*This work was supported in part by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract N00014-93-1-1212. The Institute for the Learning Sciences was founded in 1989 with the support of Andersen Consulting, part of the Arthur Andersen Worldwide Organization.

## 2. Planning knowledge

Unless one uses a simple framework such as a decision tree, generating text to meet a variety of social constraints means *planning* text. Furthermore, planning in a domain requires a specification of the acts, act preconditions, and act effects of the domain. Research on "speech acts" in the philosophy of language [Austin 1975; Vanderveken 1991] and in linguistics [Wierzbicka 1987] is an important resource for such act specifications for communication. Work in discourse processing has borrowed from that research to build both parsers and generators [Cohen & Perrault 1979; Moore & Paris 1994]. However, those discourse processing systems have generally used only speech acts closely related to **Inform** and **Request** acts and not more socially relevant acts such as **Praise**, **Thank**, and **Apologize** ([Bruce 1975] is an exception). Also, the analyses of speech act effects done in philosophy and linguistics do not extend beyond the *beliefs* ("illocutionary" effects) induced in hearers of an act. That is, the analyses do not include further effects on the emotions, relationships, or responsibilities of the persons involved. Clearly, such effects are an important part of social interactions and should be included in a model intended to be used by a planning social agent.

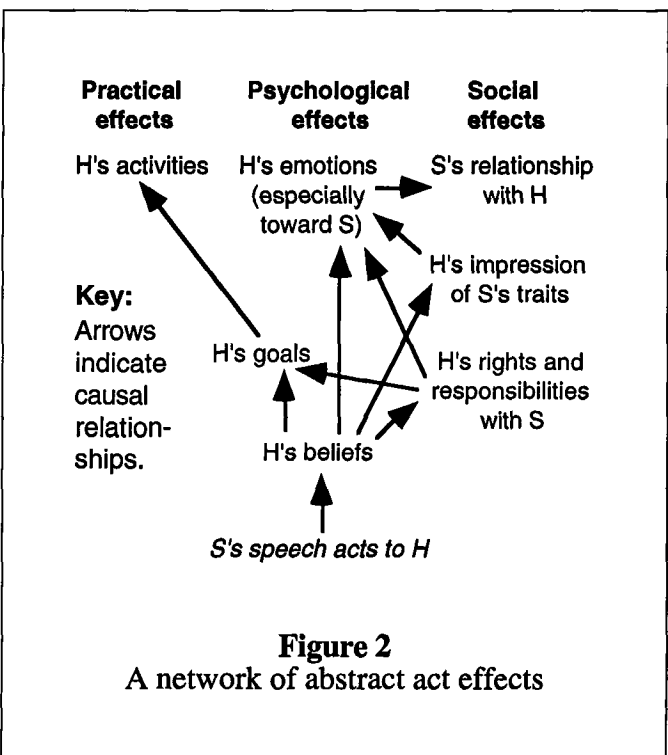
At first glance, the effects resulting from an induced belief may seem nearly innumerable. Certainly, the relations between these effects can be very complicated. But if one enumerates a small set of "effects of interest" and maps a limited set of connections, one can establish the core of a model which can be extended gradually. Figure 2 illustrates the network of abstract states which I used as a guide in developing the model. In the network, a hearer's *belief* in the content of a speech act often leads to hearer *goals*, *emotions*, beliefs about the personality *traits* of the speaker, and beliefs about the hearer's own *rights* and *responsibilities*. There are also several causal relations of interest among these later effects, such as the effect a trait impression may have on whether one likes another person. Furthermore, changes to a hearer's emotions about another person may cause changes in the hearer's attitude toward his *relationship* with that person. As an example of a chain of effects which conforms to this abstract description, consider the effects of the expression of thanks which appeared in Figure 1. The usual effect of a **Thank** act is a belief of the hearer that the speaker feels gratitude. Such a belief may then lead the hearer to have a goal to say, "You're welcome" in response. Also, the

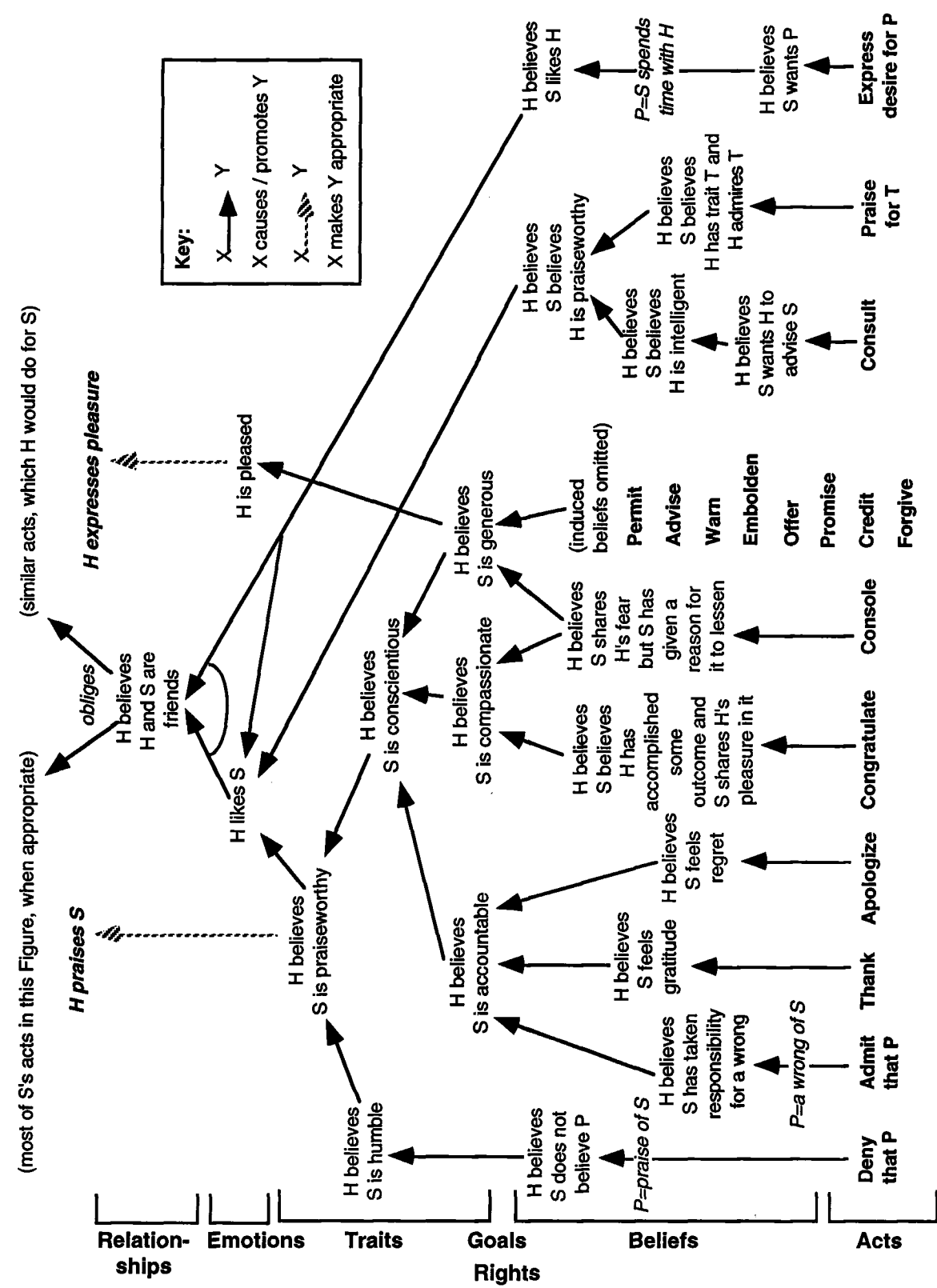
belief may give the impression that the speaker is conscientious (due to his acknowledgment of his social debt). In turn, such an impression may lead the hearer to begin to like the speaker (i.e. to feel a positive emotion toward him). If the feeling is mutual, the two individuals might enter into a cordial relationship.<sup>1</sup>

The network depicted in Figure 2 is useful for the analysis of the effects of many speech acts. Figure 3 illustrates several such analyses for acts which culminate in a positive effect on cordial relationships.<sup>2</sup> (The model has similar analyses for acts such as **Criticize** and **Threaten** which have negative or corrective effects on cordial relationships.)

<sup>1</sup>The **Thank** example does not include an effect on the hearer's rights or responsibilities. Two acts which do have such an effect are **Permit** and **Prohibit**.

<sup>2</sup>Effects on goals and rights have been omitted from Figure 3 in order to reduce graphical complexity. Chains of such effects are easily imagined for acts such as **Request** and **Permit**.





**Figure 3**  
A partial network of specific effects

Figure 3 also illustrates a pair of act preconditions (or "appropriateness conditions"). Preconditions are as important to the success of social acts as to the success of physical acts. For example, inappropriate praise may appear awkward or even sarcastic to a hearer, which would thwart any motive the speaker may have to ingratiate himself.

As in many physical domains, social act effects and act preconditions often mesh to form standard sequences of interaction, such as a speaker's praising of a hearer followed by the hearer's thanking of the speaker for his praise. These standard sequences or scripts often create expectations to which one must be sensitive as a social participant. For example, a speaker who praises a hearer but does not receive any thanks in return is likely to believe that the hearer is ungrateful. (Failing to show gratitude for an invitation causes a similar effect, due to the failure of a similar script expectation.) Avoiding such perceptions seems to be a motivation for many social actions.

An interesting feature of the model is its representation of self-reinforcing cycles of interaction. In Figure 3, cordial relationships are shown to oblige cordial acts, which links the "top" of the model to the "bottom". Similarly, unfriendly acts often lead to adversarial relationships, which may induce the participants to act on opportunities to inconvenience each other. The model represents cordial relationships as mutual liking among participants, and it represents adversarial relationships as mutual dislike. Thus, an agent can escape a cycle by indicating to others in a relationship that feelings of like or dislike are not mutual. Indicating that feelings are not mutual should be as simple as performing an act from the cycle one wishes to move into. If the other persons in the relationship reciprocate, a new type of relationship becomes active and one enters a new cycle of interaction. This description appears to reflect human relationship interactions somewhat faithfully.

The model is currently limited to two-person interactions. In the future, I intend to include inherently three-party interactions such as defaming someone and showing favoritism.

### 3. The application

The model of social acts, effects, preconditions, and scripts described above has been applied to e-mail generation via the LetterGen agent. LetterGen allows its user to

specify a high-level communicative goal (e.g. decline an invitation politely) and the agent uses its planning knowledge to suggest speech acts for an e-mail message which are appropriate for the given goal. If the user approves of a suggested act, LetterGen may then query him or her for background information to be used in the instantiation of a sentence text template associated with the speech act.

The note in Figure 1 was generated by LetterGen.<sup>3</sup> In response to an input goal to decline politely, the agent suggests seven acts: **Thank, Decline, Apologize, Make-excuse, Advise, Reassure, and Request**. An organizational template is used to place these acts in the e-mail message in the order given above.

Unlike traditional planners, LetterGen does not generate all of its suggestions in a means-end manner. Only the **Decline** act is generated in that way. The **Thank, Apologize, and Make-excuse** acts are suggested in response to the agent's fear that the **Decline** act endangers the user's goal to have the addressee like him or her. (This danger would come from the failure of the addressee's expectation that the user be grateful and polite.) The **Reassure** act is planned in a similar way, as a reaction to a "fear" that the addressee may be skeptical about the user's advice. The **Advise** act is an expression of the user's personality trait of helpfulness. This expression of helpfulness is triggered when the agent notices that the addressee needs someone to replace the user in giving a speech. LetterGen has four general plan types:

1. Goal pursuit (means-end analysis).
2. Cost avoidance, i.e. avoidance of undesired aspects of a current or incipient situation, such as unwanted social perceptions of oneself.
3. Status-quo maintenance, i.e. selection of an act because one of its effects would reinforce a desired aspect of the current situation. For example, one might offer to help another person because it would reinforce one's self-image as a generous person.

---

<sup>3</sup>The text in upper-case is text given by the user in response to a query; text in lower-case is either template text or text from rules for selecting pronouns, honorifics, and so on from other information provided by the user.

4. Trait-based habit, i.e. the performance of an act as a timeworn expression of a personality trait.

The last three plan types are triggered opportunistically.

Each plan type is represented as a rule, not a procedure, and the antecedents of the rules bind to representations of the current situation, the user's social goal, act effects, act preconditions, and other background knowledge. The consequent of these rules is a goal to perform a specific act.

One benefit of representing the types as rules is that it allows the types to be used to anticipate the plans of others. For example, the strategy of enticing someone to grant a favor by offering them something could be said to involve two nested plan types: the other person's standard plan to return favors, and one's own plan to exploit his plan by offering him something in order to get a favor in return.

If one attempts to use LetterGen's plan types to understand why a particular speech act was included in a letter, one finds that often more than one plan type could have been involved. For example, the **Thank** act might have been included in the example of Figure 1 in order to lessen the social debt the invitee owes to the inviter, or to avoid insulting the inviter through curtness, or to make the invitee feel that he is a polite person, or simply out of habit. LetterGen's model allows for all of these different plans, but in practice only one plan is used to generate an act suggestion.

LetterGen is entirely rule-based. Scripts of standard act plans are encoded as chains of rules rather than as special data structures. Forty different speech acts are defined in the agent; many of these acts have more than one sentence template associated with them, to reflect the different types of content communicated by the acts (e.g. expressing a desire to get together to celebrate versus expressing a desire to be hired). There are approximately three hundred generalized states in its social model. The agent is able to generate a dozen message types, each with many variations due to the user's ability to accept or reject act suggestions:

1. Initiate a friendship
2. Apply for a job
3. Terminate a friendship
4. Resign from a job
5. Congratulate

6. Console
7. Thank
8. Apologize
9. Accept an invitation
10. Decline an invitation
11. Encourage someone to do something
12. Discourage someone from something

LetterGen makes at least three act suggestions for each of these options. For example, the thanking option leads to a suggestion not just for **Thank** but also for **Credit** and **Offer** (to repay).

The greatest limitation of the agent is the genericness of its sentence templates, although this must be weighed against the goal to prevent the agent from becoming too dependent on the user for background information. If the agent requires too much input, it loses its appeal as a work-saving device. (Notice that this genericness is not a problem which could be solved by replacing the templates with a low-level generator.)

#### 4. A second example

In order to provide a glimpse of the variety of states represented in the model, and to illustrate further some of the limitations in LetterGen's generation ability, this section presents an example of a message aimed at terminating a friendship.

Because cordial relationships require mutual feelings of liking, the simplest way of terminating such a relationship is for one to indicate one's dislike of the other person. A less direct method is to induce the other person to dislike oneself. This effect can be induced in a number of ways, such as making oneself appear to have bad traits (e.g. intrusiveness) and giving the impression that one does not respect the other person.

Given the goal of terminating a friendship, LetterGen uses the Goal-pursuit plan type to suggest an **Express-anger** act, because it would give an impression that the writer is irascible and thereby cause the addressee to dislike him. Similarly, LetterGen suggests **Denigrate** (i.e. mention a bad trait of the hearer) because it gives an impression of pettiness, and **Prohibit** because it gives an impression of burdensomeness. **Blame** is suggested because it gives an impression of one's disrespect for the other person.

The four acts mentioned above are not "focussed" on a particular situation as the acts in the previous example were for the invitation.

I am very annoyed at YOUR BEHAVIOR AT MY PARTY THE OTHER NIGHT. THE POLICE BUST wouldn't have happened if it hadn't been for your LOUD CAR STEREO. You are so IMMATURE. You are not to COME TO ANY MORE OF MY PARTIES.

**Figure 4**  
A letter aimed at terminating a friendship

That is, at this point in the interaction, LetterGen does not know what to express anger about, what trait to denigrate, and so forth. Therefore, LetterGen is much more dependent in this case on the user to provide focussed content for the acts in the form of input text. Figure 4 illustrates an e-mail message generated from the four acts suggested plus focussed text (which LetterGen queried the user for in response to the user's approval of each suggestion).

If LetterGen could parse the user's incoming e-mail, it would be somewhat less dependent on him for background information. (The social model was designed not to be biased toward generation or parsing, with future application to parsing in mind.) But the improvement would probably not be dramatic, because much of the user's outgoing messages would concern his face-to-face interactions with others, as was the case in the second example. LetterGen should also have a memory so it can develop profiles of the user and the people he interacts with most often, because those profiles could be used to mimic the user. Furthermore, the social model should have more detail so that LetterGen can target its advice about excuses, offers, and so forth more accurately.

### 5. Related work

An important influence on LetterGen's social model is the work of Schank and Abelson on *interpersonal themes*. Themes are rules which spawn goals when triggered by certain situations. For example, "when we hear that 'John loves Mary' we can predict how John will act if Mary is threatened, if she is sick, if she is happy, if another man shows interest in her, and so on. All this information is part of the 'love' theme." [Schank & Abelson 1977, p. 139]. Schank and Abelson also emphasize the role of

communicative (MTRANS) acts as both triggers and responses to themes.

Two related social agent systems are the Affective Reasoner (AR) [Elliott 1992] and the Oz Project [Reilly & Bates 1992]. AR's agents are customers and taxi drivers in a simulated TaxiWorld. The agents construe events and react to them in the emotion-laden manner one would expect in such a domain. The Oz Project is similar to AR in its emphasis on believably emotional agents (who create dramatic situations for the enjoyment of the user through their interactions with each other). In both systems, the agents not only model emotional reactions to their own surroundings but also reason about the likely emotional reactions of other agents.<sup>4</sup> The agents represent beliefs, goals, and relationships as these relate to emotions. Yet, the agents interact only with other simulated agents, not with humans and not on behalf of humans. Such an approach has the practical benefit (but potential cost to verisimilitude) of avoiding the great variety of human interaction patterns one finds in speech acts and other forms of pragmatic communication.

### 6. Future directions

LetterGen's social model is applicable to several tasks other than text generation. Because the model is represented in terms of causal chains, it can be used to:

1. Predict the reactions of others
2. Plan based on expected reactions
3. Interpret the acts of others in terms of their likely motives
4. Diagnose errors in plans, particularly errors due to overlooked effects of an act

These functionalities are not limited solely to verbal acts, either. Physical acts such as helping, impeding, and tolerating all have an immediate effect on an observer's beliefs which is similar to the effect of speech acts on belief (because these physical acts also communicate). Because of this shared starting point, all of the states of the model which follow from belief can be used for physical, communicative interactions as well.

The openness of the model to different tasks and domains is part of LetterGen's design. One thing that it seems all future proxy agents will need is an understanding of human interaction. As LetterGen's social model is expanded and

<sup>4</sup>The models of emotion used in both systems are based on the theory presented in Ortony, Clore, and Collins 1988.

improved, it has the potential to become a resource for work on proxy agents for other applications.

### **Acknowledgments**

I am grateful to Roger Schank for his guidance and support. I would also like to thank Andrew Ortony for his helpful criticism of this project.

### **References**

- Austin, J.L. 1962. *How to do things with words*. eds., J. Urmson and M. Sbisà. Harvard University Press.
- Bruce, B. 1975. *Belief systems and language understanding*. Technical Report 2973. Bolt, Beranek, and Newman.
- Elliott, C. 1992. *The Affective Reasoner: A process model of emotions in a multi-agent system*. Technical Report 32, Institute for the Learning Sciences, Northwestern University.
- Etzioni, O., N. Lesh, and R. Segal. 1993. Building softbots for UNIX (preliminary report). Technical Report 93-09-01, University of Washington.
- Moore, J.D. and C.L. Paris. 1994. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics* (19) 4. 651-694.
- Ortony, A., G. Clore, and A. Collins. 1988. *The Cognitive Structure of Emotions*. New York: Cambridge University Press.
- Perrault, C.R. and J. Allen. 1980. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics* (6) 3-4. 167-182.
- Reilly, W. and J. Bates. 1992. *Building Emotional Agents*. Technical Report 143, Computer Science Department, Carnegie Mellon University.
- Schank, R. and R. Abelson. 1977. *Scripts, Plans, Goals, and Understanding: An inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tambe, M., W. Johnson, R. Jones, F. Koss, J. Laird, P. Rosenbloom, and K. Schwamb. 1995. Intelligent Agents for Interactive Simulation Environments. *AI Magazine* 16 (1). 15-39.
- Vanderveken, D. 1991. *Meaning and speech acts*. Volume 1. New York: Cambridge University Press.
- Wierzbicka, A. 1987. *English speech act verbs*. Sydney: Academic Press.