"How about this weather?"
Social Dialogue with Embodied Conversational Agents

Timothy Bickmore, Justine Cassell

Gesture and Narrative Language Group
MIT Media Laboratory
E15-315
20 Ames St., Cambridge MA
{bickmore, justine}@media.mit.edu

Abstract

Social dialogue can be used by an embodied interface agent both to satisfy "interpersonal" goals, such as building rapport with the user and demonstrating the expertise of the agent, and to directly or indirectly satisfy task goals, such as acquiring background information about the user. We describe the ongoing development of an embodied conversational agent that is capable of multimodal input understanding and output generation and operates in a limited application domain in which both social and task-oriented dialogue are used. An approach to discourse planning for social dialogue is described, along with an implementation and preliminary evaluation.

Introduction

People use conversational language to establish and maintain social relationships, as well as to accomplish tasks. Building rapport and common ground through small talk, intimacy through self-disclosure, credibility through "sophisticated" linguistic style, social networks through gossip, and "face" through politeness, are all examples of this phenomenon. These uses of language are not important just in purely social settings, but are also crucial to the establishment and maintenance of any collaborative relationship.

Computerized conversational agents may also profitably use language in these ways if they are to function successfully in roles which require users to interact with them for more than a few minutes, or in which we expect users to take them seriously enough to discuss their medical problems or give out their credit card numbers. Agents of this sort must be able to establish social relationships with users in order to engage their trust which, in turn, eases cooperation. In addition, language that develops social relationships heightens the illusion of agency and thus facilitates human-like interaction.

Previous research has demonstrated that people tend to respond to computers as social actors. In study after study, Reeves & Nass have shown that people respond in similar ways to computers as they do to people with respect to psychosocial phenomena such as personality, politeness, flattery, and in-group favoritism (Reeves and Nass 1996). More recent studies have demonstrated that these results also hold for some of the ways that people use language to affect their social relationships. Morkes, Kernal and Nass demonstrated that computer agents which use humor are rated as more likable, competent and cooperative than those that do not (Morkes, Kernal et al. 1998). Moon demonstrated that a computer which uses a strategy of reciprocal, deepening self-disclosure in its "conversation" with the user will cause the user to rate it as more attractive, divulge more intimate information, and become more likely to buy a product from the computer (Moon 1998).

In this paper we will focus on a conversational agent's use of small talk. Small talk is most commonly thought of as what strangers do when they meet but, in general, it can be taken as any talk in which interpersonal goals are emphasized and task goals are either non-existent or de-emphasized. Malinowski coined the term *phatic communion* to describe this kind of talk "in which ties of union are created by a mere exchange of words" (Malinowski 1923). More recent work has shown that *phaticity* is a conversational goal which is present to varying degrees in all speech (Coupland, Coupland et al. 1992).

Within task-oriented encounters, small talk can help an agent to achieve its goals by "greasing the wheels" of task talk. It can serve a transitional function, providing a ritualized way for people to move into conversation in what may be an otherwise awkward or confusing situation (Jaworski and Coupland 1999). Small talk can also serve an exploratory function by providing a conventional mechanism for users to establish the capabilities and credentials ("communal common ground" (Clark 1996)) of the agent (and vice-versa). Small talk can build solidarity with users if agents engage in a ritual of showing agreement with and appreciation of user's utterances (Malinowski 1923), (Cheepen 1988; Schneider 1988). Finally, an agent can use small talk to establish its

expertise, by relating stories of past successful problem-solving behavior, and to obtain information about the user that can be used indirectly to help achieve task goals (e.g., finding out that the user drives a minivan increases the probability that s/he has children).

One domain in which small talk is especially important is sales. Within sales, trust of the salesperson is of particular importance, especially in major purchase decisions such as real estate (Prus 1989). If the concept of trust is taken to be a composite of the perceived benevolence and credibility of an agent (Doney and Cannon 1997), then small talk should contribute directly to building trust with the user, by building rapport and establishing credibility and expertise.

REA: An Embodied Conversational Real-Estate Agent

The goal of the Rea project at the MIT Media Lab is the implementation and evaluation of an embodied, multi-modal real-time conversational interface agent. Rea implements the social, linguistic, and psychological conventions of conversation to make interactions with a computer as natural as face-to-face conversation with another person. Rea differs from other dialogue systems, and other interface agents in three ways:

- Rea has a human-like body, and uses her body in human-like ways during the conversation. That is, she uses eye gaze, body posture, hand gestures and facial displays to organize and regulate the conversation as well as to move the content of the conversation forward.
- The underlying approach to conversational understanding and generation in Rea is based on discourse functions. Thus, each of the user's inputs is interpreted in terms of its conversational function and responses are generated according to the desired function to be fulfilled (Cassell, Bickmore et al. 1999).
- Rea is able to respond to visual, audio and speech cues normally used in face to face conversation, such as speech, shifts in gaze, gesture, and non-speech feedback sounds. She is also able to generate these cues, ensuring symmetry between input and output modalities.

Rea has a fully articulated graphical body, can sense the user passively through cameras and audio input, and is capable of speech with intonation, facial display, and gestural output. Rea is displayed on a large projection screen, in front of which the user stands (see Figure 1). Two cameras mounted on top of the screen track the user's head and hand positions, while a microphone captures speech input. A single SGI Octane computer runs the graphics and conversation engine of Rea, while several other computers manage the speech recognition and generation, and image processing.

Rea simultaneously processes the organization of conversation and its content. When the user makes cues typically associated with turn taking behavior such as

Figure 1. User interacting with Rea

gesturing, Rea allows herself to be interrupted, and then takes the turn again when she is able. She is able to initiate conversational repair when she misunderstands what the user says, and can generate combined voice and gestural output. An incremental natural language generation engine based on (Stone, 1998), and extended to synthesize redundant and complementary conversational hand gestures, generates Rea's responses.

REA is an acronym for "Real Estate Agent", and within this domain we are currently focused on modeling the initial interview with a prospective buyer. We selected real estate sales specifically for the opportunity to explore a task domain in which a significant amount of social dialog normally occurs.

Discourse Planning for Social Dialog

Conversation to achieve social goals, such as small talk, places many theoretically interesting demands on dialog systems, a number of which have not been adequately – or at all -- addressed by existent approaches to discourse planning. A discourse planner for social dialog must be able to manage and pursue multiple conversational goals (Tracy and Coupland 1991), some of which may be persistent or non-discrete. For example, in small talk where there are apparently no task goals being pursued, interlocutors are conscious of multiple goals related to conversation initiation, regulation and maintenance (Cegala, Waldro et al. 1988). In primarily task-oriented interactions, speakers may also have several interpersonal goals they are pursuing, such as developing a relationship (e.g., befriending, earning trust) or establishing their reputations or expertise. It is not sufficient that a discourse planner work on one goal at a time, since a properly selected utterance can, for example, satisfy a task goal by providing information to the user while also advancing the interpersonal goals of the agent. In addition, many goals, such as intimacy or face goals (Coupland, Coupland et al. 1992) (Goffman 1983), are better represented by a model in

which degrees of satisfaction can be planned for, rather than the discrete all-or-nothing goals typically addressed in AI planners (Hanks 1994). The discourse planner must also be very reactive, since the user's responses cannot be anticipated. The agent's goals and plans may be spontaneously achieved by the user (e.g., through volunteered information) or invalidated (e.g., by the user changing his/her mind) and the planner must be able to immediately accommodate these changes.

The action selection problem (deciding what an autonomous agent should do at any point in time) for conversational agents includes choosing among behaviors with an *interactional function* such as conversation initiation, turn-taking, interruption, feedback, etc., and behaviors with a *propositional function* such as conveying information. Within computational linguistics, the dominant approach to determining appropriate propositional behaviors has been to use a speech-act-based discourse planner, on top of a classical "static world" planning mechanism like STRIPS (Fikes and Nilsson 1971), to determine the semantic content to be conveyed. Once the content is determined, other processes are typically used to map the semantic representations onto the words the agent actually speaks. Other approaches have included the use of rhetorical relations (Hovy 1988) and domain-specific knowledge (Sibun 1992) to guide the production of propositional output, but these are no more adequate to planning for social dialog.

Given the novel requirements that social discourse places upon a conversational agent's action selection mechanism, in particular the requirements for pursuing multiple, non-discrete goals in a dynamic, real-time task domain, we have moved away from static world discourse planning, and are using an activation network-based approach based on Maes' *Do the Right Thing* architecture (Maes 1989). This architecture provides the capability to transition smoothly from deliberative, planned behavior to opportunistic, reactive behavior, and is able to pursue multiple, non-discrete goals. In our implementation each node in the network represents a *joint project* (Clark 1996) -- such as telling a story, asking a question, or making a small talk contribution -- whose execution can span multiple turns. The selection of which joint project should be pursued by Rea at any given time is a non-discrete function of the following factors:

- Solidarity -- Rea continually assesses her solidarity with the user, modeled as a scalar quantity. Each conversational topic has a pre-defined, pre-requisite solidarity that must be achieved before Rea can introduce the topic. Given this, the system can plan to perform small talk in order to "grease the tracks" for task talk, especially about sensitive topics like finance.
- Topic -- Rea keeps track of the current and past conversational topics. Joint projects which stay within topic (maintain topic coherence) are given preference over those which do not. In addition, Rea can plan to execute a sequence of joint projects which gradually transition the topic from its current state to one that Rea wants to talk about (e.g., from talk about the weather, to talk about Boston weather, to talk about Boston real estate).
- Relevance -- Rea maintains a list of topics that she thinks the user knows about, and the discourse planner prefers joint projects which involve topics in this list. The list is initialized to things that anyone talking to Rea would know about--such as the weather outside, Cambridge, MIT, or the laboratory that Rea lives in.
- Task goals -- Rea has a list of prioritized goals to find out about the user's housing needs in the initial interview. Joint projects which directly work towards satisfying these goals (such as asking interview questions) are preferred.
- Logical preconditions -- Joint projects have logical preconditions (e.g., it makes no sense for Rea to ask users what their major is until she has established that they are students), and are not selected for execution until all of their preconditions are satisfied. Following Maes' approach, joint projects which enable the preconditions of other joint projects are preferred.

One advantage of the activation network approach is that by simply adjusting a few gains we can make Rea more or less coherent, more or less polite (attentive to solidarity constraints), more or less task-oriented, or more or less deliberative (vs. reactive) in her linguistic behavior.

Implementation

The discourse planner module for Rea currently manages the interview phase with users, in which Rea determines their housing needs. In this phase, the dialogue is entirely Rea-initiated, and user responses are recognized via a speaker-independent, grammar-based, continuous speech recognizer (currently IBM ViaVoice). The active grammar fragment is specified by the current joint project, and for responses to many Rea small talk moves the content of users' speech is ignored; the fact that they responded at all is enough to advance the dialogue.

Joint projects are specified by:

- Type -- e.g., small talk statement, small talk query, task statement, task query
- Rea moves -- currently encoded as surface strings
- User moves -- speech recognition grammar specification and whether a user response is required at all
- Conversational topics covered
- Logical preconditions and whether the joint project contributes to a goal, as described above

At each step in the conversation in which Rea has the floor (as tracked by a conversational state machine in Rea's Reaction Module), the discourse planner is consulted for the next joint project to initiate. At this point, activation values are incrementally propagated through the network (following (Maes 1989)) until a joint project is selected whose preconditions are satisfied and whose activation value is over a specified threshold.

Topic shifts are marked by discourse markers and beat gestures. Discourse markers include "so" on the first small

talk to task talk transition, "anyway" on resumption of task talk from small talk, and "you know" on transition to small talk from task talk.

Within this framework, Rea decides to do small talk whenever solidarity with the user needs to be increased (e.g., before a task query can be asked), or the topic needs to be moved incrementally to a desired topic and small talk contributions exist which can facilitate this. The activation energy from the user relevance condition described above leads to Rea starting small talk with topics that are known to be in the shared environment with the user (e.g., talk about the weather or the lab).

Example Interactions

An interview between Rea and a user typically proceeds as shown in the following dialogue (baseline case). (User responses are only shown in positions in which they effect the selection of subsequent joint projects.)

1. That microphone is terrible, I hate using those things.
2. Sorry about my voice, this is some engineer's idea of natural sounding.
3. Are you one of our sponsors? *User: Yes*
4. Were you at our last sponsor meetings?
5. I got so exhausted at the last sponsor meeting I think I was starting to lose my voice by the end.
6. So, where would you like to live?
7. How many bedrooms do you need?
8. Do you need access to the subway?
9. Is one bath enough?
10. You know, Boston is certainly more expensive than it used to be.
11. Anyway, what can you afford?
12. What kind of down payment can you make?
13. Let me see what I have available.

Example 1. "Normal Rea"

In this example, Rea opens with small talk moves regarding things in her shared physical environment with the user (1-2). She then proceeds to small talk related to sponsors (after establishing that the user is a sponsor). After a few turns, enough solidarity has been established (simply by doing small talk) that Rea can move into task talk (6-9). However, before bringing up the topic of finance--a topic that is potentially very face threatening for the user--Rea decides that additional solidarity needs to be established, and moves back into small talk (10). This small talk move not only increases solidarity but shifts the topic to finance, enabling Rea to then bring up the issue of how much the user is able to afford (11-12).

If Rea's adherence to solidarity preconditions is reduced, by decreasing the contributions of these preconditions to the activation of joint projects, this results in her being less polite and empathetic, and more task goal oriented. If everything else is held constant (relative to the baseline example) the following dialogue is produced.

1. So, where would you like to live?
2. What can you afford?
3. What kind of down payment can you make?
4. How many bedrooms do you need?
5. Do you need access to the subway?
6. Is one bath enough?
7. Let me see what I have available.

Example 2. "Impolite Rea"

In this example, Rea does not perform any small talk and sequences the task questions in strictly decreasing order of priority.

As a final example, the importance of task goal satisfaction can be reduced (relative to the baseline case) yielding the following dialogue in which Rea rambles on at length before getting down to business.

1. That microphone is terrible. I hate using those things.
2. Sorry about my voice. This is some engineer's idea of natural sounding.
3. Are you tired of demos yet?
4. Are you one of our sponsors? *User: Yes*
5. Were you at our last sponsor meetings?
6. I got so exhausted at the last sponsor meeting. I think I was starting to lose my voice by the end.
7. I have lived in Boston all my life. Come to think of it, I have lived inside this room all of my life. It is so depressing.
8. How about this weather?
9. I think winters in Boston are awful
10. Boston is certainly more expensive than it used to be.
11. So, what can you afford?
12. What kind of down payment can you make?
13. You know, I keep showing the researchers here the same houses, over and over again. Maybe one day I will get lucky.
14. Have you been in the Media Lab before?
15. It is pretty cool do you think?
16. They are doing some crazy things in here.
17. Do you know that the Media Lab is going to expand into another building. Things are really going well for the researchers here.
18. I really love it here at M I T. Do you?
19. It is pretty intense, do you think?
20. I have shown houses to lots of students and faculty from M I T. But I always enjoy talking to them.
21. Anyway, where would you like to live?
22. How many bedrooms do you need?
23. Do you need access to the subway?
24. Is one bath enough?
25. Let me see what I have available.

Example 3. "Chatty Rea"

When task goals become less important, topic coherence and other source of activation energy become the primary factors in deciding what Rea says next. In this example, by

the time Rea has transitioned into the topic of finance enough solidarity has been established that she can immediately ask the user how much they can afford (11-12).

Evaluation

We are currently conducting an experiment to assess the efficacy of small talk in real estate interviews with Rea, and its effect on the user's perception of and trust in Rea. Subjects are interviewed by Rea about their housing needs, shown two "virtual" apartments, and then asked to fill out a questionnaire, which includes questions about how much rent they would be willing to pay for one of the apartments, attitudinal measures of perceived likability, competence, and intelligence of Rea, and a standard measure of trust (Wheeless and Grotz 1977). This is a between-subjects experiment with two conditions: TASK - Rea does not do any small talk; and NATURAL - Rea does a "natural" amount of small talk representative of the output of the discourse planner described above. To avoid possibly confounding conditions, no humorous or self-disclosing small talk utterances are used, and Rea is controlled using a wizard-of-oz configuration so that variations in speech recognition accuracy do not adversely affect the results. A pilot study indicated that users trust Rea more and like her more in the NATURAL condition (example given above), but significant results await completion of the full study.

Conclusion

Social intelligence includes knowledge of when and how to use language to achieve social goals. This knowledge is crucial for our computational agents if they are to be as effective as people, and if we want people to be able to use our agents easily, efficiently, and cooperatively. As embodied conversational agents become ubiquitous, the ability for them to establish and maintain social relationships with us will become increasingly important. It is possible to specify social and task goals in such a way that small talk can be generated to achieve social goals such as decreased interpersonal distance, which in turn achieves task goals such as achieving a sale.

References

Cassell, J., T. Bickmore, et al. (1999). *Embodiment in Conversational Interfaces: Rea*. CHI 99, Pittsburgh, PA.

Cegala, D., V. Waldro, et al. (1988). *A study of interactants' thoughts and feelings during conversation.* Ninth Annual Conference on Discourse Analysis, Philadelphia, PA.

Cheepen, C. (1988). *The Predictability of Informal Conversation.* New York, Pinter.

Clark, H. H. (1996). *Using Language*. Cambridge, Cambridge University Press.

Coupland, J., N. Coupland, et al. (1992). ""How are you?": Negotiating phatic communion." *Language in Society* 21: 207-230.

Doney, P. and J. Cannon (1997). "An Examination of the Nature of Trust in Buyer-Seller Relationships." *Journal of Marketing* 61: 35-51.

Fikes, R. and N. Nilsson (1971). "STRIPS: A new approach to the application of theorem proving to problem solving." *Artificial Intelligence* 5(2): 189-208.

Goffman, E. (1983). *Forms of Talk*. Philadelphia, PA, University of Pennsylvania Publications.

Hanks, S. (1994). *Discourse Planning: Technical Challenges for the Planning Community.* AAAI Workshop on Planning for Inter-Agent Communication.

Hovy, E. H. (1988). *Planning Coherent Multisentential Text*. ACL, Buffalo, NY.

Jaworski, A. and N. Coupland (1999). *The Discourse Reader*. London, Routledge.

Maes, P. (1989). "How to do the right thing." *Connection Science Journal* 1(3).

Malinowski, B. (1923). The problem of meaning in primitive languages. *The Meaning of Meaning.* C. K. Ogden and I. A. Richards, Routledge & Kegan Paul.

Moon, Y. (1998). Intimate self-disclosure exhanges: Using computers to build reciprocal relationships with consumers. Cambridge, MA, Harvard Business School.

Morkes, J., H. Kernal, et al. (1998). *Humor in Task-Oriented Computer-Mediated Communication and Human-Computer Interaction*. CHI 98.

Prus, R. (1989). *Making Sales: Influence as Interpersonal Accomplishment*. Newbury Park, CA, Sage.

Reeves, B. and C. Nass (1996). *The Media Equation: how people treat computers, televisions and new media like real people and places*. Cambridge, Cambridge University Press.

Schneider, K. P. (1988). *Small Talk: Analysing Phatic Discourse*. Marburg, Hitzeroth.

Sibun, P. (1992). "Generating Text Without Trees." *Computational Intelligence: Special Issue on Natural Language Generation* 8(1): 102-122.

Tracy, K. and N. Coupland (1991). Multiple goals in discourse: An overview of issues. *Multiple goals in discourse*. K. Tracy and N. Coupland. Clevedon, Multilingual Matters**:** 1-13.

Wheeless, L. and J. Grotz (1977). "The Measurement of Trust and Its Relationship to Self-Disclosure." *Human Communication Research* 3(3): 250-257.