Dialogue Generation in an Assistive Conversation Skills Training System for Caregivers of Persons with Alzheimer's Disease

Nancy L. Green
Department of Mathematical Sciences
University of North Carolina at Greensboro
Greensboro, NC 27402 USA
nlgreen@uncg.edu

Boyd Davis
Department of English
University of North Carolina at Charlotte
Charlotte, NC 28223 USA
bdavis@email.uncc.edu

Abstract
We describe a proposed assistive conversational skills training system using artificial intelligence and natural language generation techniques to simulate spoken dialogue between two embodied agents, one character representing a caregiver and the other a person with Alzheimer's Disease (AD). The type of dialogue simulated is social conversation in which the caregiver encourages the storyteller to tell autobiographical stories. Intervention by the caregiver may be required at times to keep the conversation going due to linguistic or cognitive problems experienced by the storyteller. In our proposal, turn and topic management behavior for the storyteller character is implemented by combining rules representing normal pragmatic routines with linguistically-motivated rules representing coping strategies.

Introduction
The goal of our research is to develop an interactive virtual world for training caregivers to participate in sustained social conversations with persons who have Alzheimer's Disease (AD) (Green 2002). Cognitive-linguistic stimulation can improve or maintain the functioning of Alzheimer's dementia patients (Mahendra 2001). Gerontological research has discussed the psychological benefits to the elderly of recounting their autobiographical memories (Mills and Coleman 1994; Golander and Ras 1996). However, although well-meaning, without training caregivers find it frustrating and difficult to engage in social conversation with a person with AD; furthermore, they may not recognize attempts by a person with AD to share reminiscences or to engage in dialogue for social purposes (Small et al. 1998; Davis, Moore and Brewer 2001).

Our proposed system will coach caregivers on assistive conversational techniques by enabling the caregiver to converse, through his or her virtual world avatar (henceforth referred to as "the Avatar"), with another character simulating a conversational partner who has AD (henceforth referred to as "the ADC"). Artificial intelligence and natural language generation techniques will be used to generate the linguistic and other behaviors of these animated characters. At pedagogically significant points in the dialogue, the user (i.e. caregiver) will be allowed to select the Avatar's next dialogue action from a set of choices, see and hear the action being performed, and finally observe its effect on the ADC and the simulated conversation. We hope that, by having the opportunity to practice assistive dialogue techniques in a realistic simulation environment, the user will acquire techniques that will transfer to his or her conversational interactions with real persons in his or her care. By realistic we mean that the dialogue should resemble naturally occurring social conversations with the elderly (whether affected by AD or not) about their autobiographical memories and that the dialogue should be spoken by embodied agents (i.e. characters capable of speech and of conveying emotion through their bodily actions). Figure 1 gives an excerpt from a naturally occurring dialogue of this sort (Shenk et al. 2002).

The potential impact of AD on conversational behavior can be seen by considering the CAPPCI inventory, which was designed for evaluating linguistic deficits of persons with cognitive impairments (Perkins et al. 1997). Features that are inventoried include: turn-taking skills (e.g., the ability to initiate conversation, to respond when selected as next speaker, or to hand over conversational floor), topic management skills (e.g., ability to initiate new topics and to maintain topics), memory (e.g., failure to remember family or events discussed in conversation), and linguistic abilities (e.g., failure in word retrieval, overuse of pronouns, production of circumlocutions). For example in the excerpt in Figure 1, Glory (a speaker with AD) has difficulty listing all the items that were grown on her childhood farm. On the other hand, she tells an entertaining and coherent story about herself. From a
In fact, while much research on the impact of AD on communication has focused on describing the associated deficits, recent research has suggested that Alzheimer's speakers retain some pragmatic competence that they can use to maintain conversational fluency (Davis, Moore, & Peacock 2000; Davis, Moore, & Brewer 2001; Davis, Moore, & Greene 2002). In addition, these researchers are studying assistive conversational techniques that a caregiver can use to help sustain a social conversation with an Alzheimer's speaker and to help "co-construct" (Jacoby & Ochs 1995) the latter's reminiscences. These techniques are illustrated by the contributions of the speakers labelled LM and BD in the excerpt in Figure 1. Shenk et al. (2002) found that co-constructed stories of elderly persons with AD are similar in many respects to stories told by elderly persons without AD. The results of research on these assistive techniques (e.g. Moore & Davis 2002) will contribute to the repertoire of interventions on which our proposed system will provide coaching.

In this paper we describe our proposed approach to the generation of discourse actions in a two-person social conversation by the participant who is being encouraged to tell an autobiographical story (e.g. an elderly person) by the other participant (e.g. a caregiver). We focus on the problems of turn management, topic regulation, and fulfilling discourse obligations within the time scale of naturally occurring dialogue. Our proposed approach is to provide a computational model that integrates normal pragmatic functioning with coping strategies for problems due to cognitive constraints such as those experienced by persons with AD. This approach is motivated in part by the hypothesis of the above researchers that pragmatic routines are retained and used by speakers with AD even in later stages of the disease. Another motivation for this approach is that a unified model may be useful for generating realistic conversational behaviors in systems simulating dialogue between normal speakers, since even normal speakers at times are subject to heavy cognitive loads, distractions, or other problems that interfere with linguistic or memory processes. Finally, such a model may simplify the simulation of characters representing speakers at different stages of AD or manifesting other individual differences in the effects of AD. The proposed system described in this paper has not been implemented yet.

In the rest of the paper, we first summarize the high-level design of the proposed system. Then we describe relevant research on AD discourse, followed by our proposal for discourse action generation. Next, we discuss related work in dialogue generation. Lastly, we discuss our plans for future work including evaluation of the proposed system.

**System Design**

As background, here we briefly summarize the proposed design. For more information see (Green 2002). The user interface to the proposed system will present a virtual world containing two main characters, the (caregiver's) Avatar and the ADC. These two characters engage in spoken dialogue with facial gestures. At pedagogically significant points in the dialogue, the user is offered a choice of interventions that the Avatar can make.

The internal design of the system includes two intelligent agents for generating the possible behaviors of these two characters. The Caregiver-coach Agent, which controls the Avatar, is composed of three modules. The Action Proposal Module proposes a set of possible actions that are appropriate with respect to the current Dialogue Model (described below). Next the Action Selection Module decides on pedagogical grounds whether to let the user select the next action; if the user is not granted the choice, then this module selects one. After an action has been chosen, the Dialogue Realization Module determines how to perform the action.

The Conversational Storyteller Agent, which controls the ADC, also is composed of three modules. Initially, its Action-Proposal-Selection Module selects an action from the set of actions that are appropriate with respect to the current Dialogue Model. If the selected action is to pursue a narrative goal, then the action is refined by the Storytelling Action Refinement Module. This module must extract information from the agent's Life Episodes Model, a representation of the agent's supposed autobiographical memories. After an action has been chosen, the Dialogue Realization Module determines how to perform the action. In order to generate a believable simulation of a speaker with AD, the agent will be subject to problems in each of these three modules of generation. For example, in Action-Proposal-Selection, the agent's willingness to continue telling a story may be influenced by the state of its Emotion-Personality Model: after experiencing a loss of face in the preceding discourse, the agent may no longer be willing to continue telling a story. Second, the output of the Storytelling Module may be influenced by the agent's inability to access parts of the Life Episodes Model. Third, the Dialogue Realization Module may suffer from impairment in functions such as lexical choice or referring expression generation.

The Dialogue Model includes a Discourse Model, an Emotion-Personality Model representing the affective state of the Storyteller Agent, and a Narrative Model representing the structure and content of the story that has
been told so far. Previously, we proposed that both agents would share a Dialogue Model so that the system would not need to perform interpretation; whenever an agent generated a communicative action, it would update the Dialogue Model with the intended interpretation (Green 2002). That approach may be sufficient for simulating conversation between two idealized speakers, e.g. in an interactive entertainment application. However, we have realized that a problem that our system design must address is the representation of the effects of dialogue contributions that are not understood, at least not right away, by the hearer.

To illustrate the problem, Figure 2 gives an excerpt of an actual conversation between Libby (LD), a speaker with AD, and her visitor (BD). If our system were generating this dialogue, it would be reasonable to suppose that some of LD's contributions would not be completely understood by the caregiver using our system for training (e.g. Well, I'll have to tell my husband about that because he's given me lots of decent things into rocks without asking.). The system must track when the ADC’s contribution is problematic so that the caregiver can be trained to respond appropriately. While in an information-seeking dialogue an appropriate response might be to seek clarification (What do you mean by "into rocks"?) that would undoubtedly have a detrimental effect on a conversation with an AD speaker. However, as BD's response shows, it is possible to respond in a way that keeps the conversation going. Furthermore, to model this excerpt it is not sufficient to ignore problems in comprehension, as shown by BD's response to Because I've seen several of them right in this room!; BD guesses that them refers to people from different parts of the country, which is confirmed by LD's Uh huh. Also, this excerpt shows that at times even the "normal" speaker may not be understood (How canned he is?).

**Corpus Study**

In this section we summarize results of qualitative analyses of pragmatic features of AD discourse reported in (Davis, Moore, & Peacock 2000). Based on transcripts of recordings of 75 hours of conversations between "normal" speakers (members of the research group performing the analyses) and between nine and twelve nursing home residents with AD, the analyses focused on topic-regulation.

"One problem that some moderately to severely demented AD speakers face is that they may not be quite sure any more … how to bring a topic to a close" (Davis, Moore, & Peacock 2000). Analysis of the transcripts revealed that the AD speaker often tended to use 'frozen' phrases, "phrases using colloquial, figurative, or aphoristic language … like a cliché or common saying", to solicit topic closure. In many cases, the normal partner responded by initiating an affirm-confirm-extend (ACE) routine. The ACE routine has been observed in normal conversation as a common way of closing a topic (Drew and Holt 1998). The routine begins with one participant "summarizing or evaluating [the current topic] with figurative language, frequently using a metaphor or aphorism. The second participant affirms or confirms the underlying concept, which is, in turn, reaffirmed or reconfirmed by its originator. The topic is now closed and either participant may start a new one" (Davis, Moore & Peacock 2000). Here is an example (where NS denotes the normal speaker and AS the Alzheimer's speaker):

NS: How have you been -- feeling okay?
AS: Yeah. I'm improving right along. [FROZEN]
NS: That's great! [AFFIRM]
AS: Sure is. [CONFIRM]

In the speech of one of the AD conversants, who had been classified with moderate to severe cognitive decline by his caregivers, in about 70% of the cases where he used a frozen phrase, the next turn was "followed by a turn that either starts up the ACE routine or introduces a topic shift"; furthermore, he "participates in the ACE Routine and occasionally introduces it; after the Routine, he is nearly as likely as his partner to introduce the new topic" (Davis, Moore & Peacock 2000).

In addition to their role in topic closure, frozen phrases also appear to be used by AD speakers to maintain "conversational credibility" when they are unable to answer a question; this usage is referred to as 'Frozen-Coping' (Davis, Moore & Peacock 2000). A frozen-coping phrase is usually preceded by a discourse marker and a pause and occurs at the beginning of a turn. Here is an example:

NS: Well, that's true in any job. What did you like doing?
AS: Just… what I was doing. [FROZEN COPE]
NS: Yeah.
AS: At the center -- at Southern Furniture

Although from a normal speaker a response such as that given by AS to the preceding question might be interpreted as evasive or vague, in this excerpt NS provides positive feedback, enabling AS to contribute what he can in the next turn without losing face.

**Generating Discourse Actions**

In our proposed model, the Action-Proposal-Selection and Dialogue Realization modules of the Conversational Storyteller Agent use reactive, condition-action rules for turn-management, topic-regulation, and realization of discourse markers and formulaic phrases. We have
encoded about 20 rules describing "normal" dialogue generation. To paraphrase a representative sample:

- when agent A is offered the turn by conversational partner and A accepts, A forms an interactional goal to signal acceptance of the turn,
- talking signals acceptance of the turn,
- talking holds the turn (when an agent pauses mid-turn, if he does not resume talking within a certain amount of time, he may lose the turn, or at least lose face),
- if the partner asked agent A a question in the preceding turn and A has the current turn, then A has a discourse obligation to answer the question,
- if agent A holds the turn and has no discourse obligations and wants to change the current global topic T, then A forms a content-selection goal to close T,
- if agent A holds the turn and has a discourse obligation to answer a question Q, then A forms a content-selection goal to answer Q,
- if agent A finds speech act S and content C to satisfy his content-selection goal G, then A forms an utterance-planning goal to realize <S,C,G>.
- if agent A has planned utterance U to realize <S,C,G> then A utters U and receives credit for U's goodness and discharges the associated discourse obligation.

To account for the use of frozen phrases and discourse markers in the speech of AD speakers, we have added several rules that reflect the interpretation suggested by Davis, Moore and Peacock (2000) and also that reflect points in generation, as encoded in our "normal" rule set, where a speaker might encounter problems. To paraphrase some rules describing AD discourse:

- [Turn-initial discourse marker followed by pause] If agent A has an interactional goal to signal acceptance of the turn and the time limit within which A must start talking has almost expired and A has not planned any utterance but A can access an appropriate discourse marker D, then A utters D thereby signalling acceptance of the turn and incrementing the time limit by which time A must begin speaking again,
- [Frozen-coping] If agent A is (still) trying to satisfy a content-selection goal G to answer a question and the time limit within which A must start talking has almost expired and A has already uttered a turn-initial discourse marker to hold the turn, but A can access a frozen phrase U with content C and of a speech act type S, where S is appropriate and C is relevant to G, then A utters U and releases the turn,
- [Frozen phrase to solicit ACE] If agent A wants to change topic but has not yet satisfied his content-selection goal G to close the current topic T and the time limit within which A must start talking has almost expired, but A can access a frozen phrase U with content C and of a speech act type S, where S is appropriate and C is relevant to T, then A utters U and releases the turn.

In addition to rules for the routine behaviors encoded in the above two sets of rules, the complete model also must provide a mechanism for satisfying the content-selection and utterance-planning goals referred to in the above rules. Thus, we are considering using an architecture like that used in (Green and Lehman 2002) which supports integrated reactive and deliberative processing. Also, there are a number of lower-level natural language generation issues that we will need to address. First, lexical choice must reflect dialectal and other types of sociolinguistic variation. For example, in the ACE exchange given above, AS uses regional expressions such as improving right along and sure is. Use of such expressions by our system will help create vivid characters whose speech is consistent with their Life Episodes Models. Moreover, the system must track use of such expressions because of their importance in social interaction; speakers may use such expressions to create solidarity and thus, the Caregiver Avatar must be able to respond appropriately. For example, in the dialogue excerpt in Figure 1, BD responds to Glory's pickin' a hundred pounds of cotton ... UhhmmHmm with an expression outside of BD's usual dialect: An' you so tiny. Second, lower-level generation processes must perform at the rate of natural conversation. Moreover, they must "know when to quit" so that if the current goal has not succeeded within a specified amount of time, the agent's coping rules can be applied.

**Related Work**

Natural language generation by embodied conversational agents for business, entertainment, and educational applications is an active research area (e.g., Cassell et al. 1999). This area includes research on developing interactive drama systems for training purposes. Carmen's Bright Ideas is a prototype instructional system designed to allow mothers of pediatric cancer patients to improve problem-solving skills by participating, through an avatar, in a simulated counseling session (Marsella et al. 2000). As in our planned system, the user has influence over the simulated dialogue through her avatar; such user influence requires system designers to address the issue of controlling plot. Carmen's Bright Ideas uses a branching plot written by human authors and designed to advance the dialogue through the stages of the Bright IDEAS problem-solving method. In our system, we plan to use AI techniques to create the ADC's story dynamically as a function of the discourse state, the ADC's affective state, and the ADC's Life Episode Model. The Mission Rehearsal Exercise (MRE) is a prototype system designed to teach decision-making skills through direct user
participation in a virtual military mission (Swartout et al. 2001). The user plays the role of a character in the mission and communicates through spoken dialogue with the other characters. Building on previous work in AI on emotion and personality, the behavior of the MRE's artificial characters is influenced by emotion models (Gratch and Marsella 2001). In our proposed system, the ADC will exhibit the influence of the Storyteller agent's Emotion-Personality model. The model will be affected both by the discursive actions of the caregiver's Avatar and by the agent's self-perceived success in participating in the conversation.

Other work in embodied conversational agents related to ours is on generation of small talk and conversational storytelling. Bickmore & Cassell (1999; 2000) implemented a prototype embodied interface agent, REA, capable of mixing small talk and storytelling with task-oriented dialogue. A simulated real estate agent, REA employs these two strategies to keep the conversation going, build rapport with users, and thereby gain their cooperation. In addition, storytelling can be used by REA to provide information indirectly. In contrast, the ADC in our system is encouraged by the Avatar to engage in small talk and autobiographical reminiscing to provide cognitive and linguistic stimulation to the ADC and mutual enjoyment to both (imaginary) parties. Despite these differences, our system faces similar implementation requirements to those noted by Bickmore and Cassell (1999): the ability to pursue simultaneous multiple goals, the ability to pursue non-discrete interpersonal goals, and the ability to integrate reactive and deliberative behavior. REA's dialogue management module, which is responsible for sentence planning with co-verbal gesture, tracks dynamic contextual features including attentional prominence, cognitive status (hearer-old or hearer-new, and in-focus or not), and information structure (theme-rheme) (Cassell, Stone, & Yan 2000). In order for the Caregiver-coach Agent to respond with appropriate interventions for linguistic problems exhibited by the ADC, our system may have to track these features as well.

The time-sensitive turn and topic management rules discussed in this paper are related to previous work in computational models of turn-management. Aist and Mostow (1997) implemented a Reading Tutor agent that listens to children reading aloud. The Tutor generates interventions (such as speaking a word or reading an entire sentence or filling a pause with a backchannel) based upon patterns detected in the student's reading (such as a period of silence or hesitating on a difficult word). The Tutor's turn-taking rules may include a delay until a specified amount of time has elapsed since the event that triggered a rule occurred, such as the onset of the student's turn. In this paper we have discussed how time constraints of dialogue may result in the use of coping strategies such as frozen phrases by the Storytelling Agent. In addition, the Caregiver Agent in our system will need a time-sensitive mechanism for deciding when to step in during the Storytelling Agent's turn and offer an intervention. As in the Reading Tutor, some interventions may be advisable, without taking the floor, when the other speaker is having difficulty but, with appropriate help, could continue his turn.

Donaldson and Cohen (1997) provide a three-step model of turn-taking for a conversational agent. First, an agent's time-bounded persistent goals arise, e.g., to seek clarification, answer a question, or repair an apparent misunderstanding. (Time-bounded means that the agent ceases to pursue goals that may no longer be relevant due to the time that has elapsed since they were formed.) Second, these goals are ordered using a constraint satisfaction (CSP) framework. Third, after the agent has chosen which goal to attempt next, it waits until it detects a signal indicating that the partner is willing to give up the floor. Their model was proposed for task-oriented dialogue, and it is not clear whether the full power of CSP is required for social conversation.

In the Virtual Theater (Hayes-Roth & van Gent 1996), synthetic actors improvise their behavior, including dialogue with other synthetic actors and avatars responding to directives from users. Rousseau and Moulin (1997) implemented a finite-state model of conversational protocol for determining when a character in the Virtual Theater takes the turn and when it releases it. The protocol they implement is similar to our rules for "normal" conversation but does not appear to be time-sensitive.

**Future Work**

Future plans include additional analyses of the corpus, implementation of a prototype system, and evaluation. From the corpus, we wish to gain further insight into features of an AD speaker's discourse that our system must simulate and that could trigger interventions by the Caregiver-coach Agent. Also, we wish to study the narrative properties of the stories told by the AD speakers. This knowledge, coupled with knowledge gained from our colleagues' evaluations of the efficacy of various interventions is required for implementing the prototype. Currently, we are investigating available products for use in implementing the prototype's interface agents, i.e., the Caregiver Avatar and the ADC. After the prototype is implemented, it will be evaluated in several ways. One dimension is dialogue realism, i.e., the accuracy of the simulation of discourse of persons with AD. A second dimension is the efficacy of the prototype as a training tool for caregivers. The final evaluation on this dimension will consist of formal field trials comparing the communicative effectiveness of caregivers trained by our system as compared to those who have had no training.
Conclusion

We have described a proposed assistive conversational skills training system using artificial intelligence and natural language generation techniques to simulate spoken dialogue between two embodied agents, one character representing a caregiver and the other character a person with Alzheimer's Disease. The type of dialogue simulated is social conversation in which the caregiver encourages the person with AD to tell autobiographical stories. Intervention by the caregiver character may be required at times to keep the conversation going due to linguistic or cognitive problems experienced by the storyteller character. In our proposal, turn and topic management behavior for the storyteller character is implemented by combining rules representing normal pragmatic routines with linguistically-motivated rules representing coping strategies.

Acknowledgments

We wish to thank our collaborators in Nursing and Gerontology at the University of North Carolina at Charlotte for their contributions to linguistic research on discourse of Alzheimer's speakers.

References


Glory: I just lived in a regular farm home. Farmed cotton, corn, eh-everything you … grow on a farm.
BD: That's right.
Glory: I had a big ol' cotton bag tied around me, pickin' a hundred pounds of cotton … UhhmmHmm.
BD: A hundred pounds? An' you so tiny!
Glory: Huh?
LM: You're a tiny person to be carrying that much cotton.
Glory: I decided one day I'd pick a hundred pounds. Guess how much?
LM: How much?
Glory: A hundred and three.
LM: Oooohh.
BD: Wow.
Glory: I went over.
BD: That's fantastic.
Glory: A hundred and three -- you've got to grab it to … get a hundred and three pounds of cotton in one day.

Figure 1: Transcript of conversation between Glory, an elderly nursing home resident with Alzheimer's Disease and two visitors (LM and BD) from (Shenk et al. 2002).
LD: Well, I'll have to tell my husband about that because he's given me lots of decent things into rocks without asking.
BD: Everytime you have told me about your husband you have mentioned how kind he is and how generous he is.
LD: How canned he is?
BD: How kind. I know. I didn't sound right! I sounded too southern!
LD: (laughs)
BD: You New York woman!
(They both laugh.)
LD: That's very funny.
(They all laugh.)
BD: You came from New York to South Carolina, I believe, didn't you?
LD: I certainly did. Is that where you came from?
BD: No. I came from Kentucky. Which is why my accent is very strange for here.
LD: All the way through, I bet you it is.
BD: Yes. All the way through.
LD: Because I've seen several of them right in this room!
BD: I'm not surprised. There are several people from different parts of the country here.
LD: Uh huh.
BD: And each will sound different.
LD: (laughs)

Figure 2: Transcript of conversation between Libby (LD), an elderly nursing home resident with Alzheimer's Disease and a visitor (BD).