Controlling interaction with meta-acts

David Novick

Department of Computer Science and Engineering
Oregon Graduate Institute of Science and Technology
19600 NW Von Neumann Drive
Beaverton, Oregon 97006-1999
novick@cse.ogi.edu
(503) 690-1156

January 15, 1991

Abstract

User models that are adequate for conversational interaction between human and machine must account for more than the state of the user's beliefs about the domain. The model must also include meta-knowledge about the state of conversation control. This paper proposes a conversational model based on *meta-locutionary acts*, and presents a conversational simulation based on the model. Results of the simulation suggest that the meta-locutionary model appears adequate for representation for control of conversational interaction with users. For the a domain involving letter sequences, simple conversations have been simulated successfully.

Introduction

User models that are adequate for conversational interaction between human and machine must account for more than the state of the user's beliefs about the domain. The model must also include meta-knowledge about the state of conversation control. This paper discusses the representation and use of such meta-knowledge, proposes a conversational model based on *meta-locutionary acts*, and presents a conversational simulation based on the model.

Human-computer interaction tends to be fragile and frustrating while human-human interaction displays robustness over a wide variety of circumstances. Figure 1 summarizes some of the differences between the two forms of interaction.

	H H I	C H I
control	mixed-initiative	single-initiative
actions	situated	pre-planned
language	informal, relaxed	formal, stilted

Figure 1. Differences between interaction modes. Human-human interaction (HHI) is generally more flexible than human-computer interaction (HCI).

In human-computer interaction, typically either the human or the computer controls the flow and structure of the interaction. This is single-initiative interaction, in the sense that only one of the parties is empowered to initiate exchanges. This form of interaction is largely pre-planned, either by the human or by the designers of the program. Even so, it is largely because they are forced by the interface (or have learned from earlier frustration) to use well-formed language that the human and the computer are able to understand each other at all.

In contrast, human-human conversations are characterized by unpredictability, ungrammatical utterances, non-verbal expression, and mixed-initiative control in which the conversants take independent actions. In the face of an unexpected shift of conversational context, somehow humans are able to understand that the shift has occurred. Our every-day language is notoriously ill-formed, full of improper usages, mismatched agreements, halts and starts, in-line corrections of self and other, and utterances that may not even be words. Moreover, we interpret and express a wide range of non-verbal behaviors as a concomitant of verbal interaction. In human-human conversation, both conversants have the power to seize the initiative to meet their own needs and expectations. If the language of human-computer interaction seems stilted and tame, the language of ordinary human-human conversation is wild and untamed.

Meta-locutionary acts

The evidence is strong for the proposition that conversants in interactive discourse share a model of their conversation (Clark & Marshall, 1981). In informal terms, a weak version of this conjecture would be that the conversants must have at least some knowledge necessary to the conversation which is common to all conversants. A strong version is that the conversants are jointly creating a single (though possibly complex) intellectual product. Suchman (1987) characterizes conversation as an “ensemble” work:

> Closer analyses of face-to-face communication indicate that conversation is not so much an alternating series of actions and reactions between individuals as it is a joint action accomplished through the participants’ continuous engagement in speaking and listening [references omitted]. (Suchman, 1987, p. 71)

Traditional interfaces—even those based on speech acts—are largely unable to handle these aspects of “feral” language because they mainly rely on parsing sentence-level interaction and fail to account properly for the context created by the conversation itself. Conversational processes are apparently not characterized by top-down planning; they are made up of actions which are responsive to the dynamic interactive situation. The importance of feedback to conversational coherence was shown by Kraut, et al. (1982). Turn-taking (Sacks, Schegloff, & Jefferson, 1974; Duncan, 1980) is a control-oriented account of conversation that explicates patterns of mixed-initiative interaction in terms of conversational turns. This represents an organizational substrate for domain-level, intentionality-based interaction. It is maintained through a wide set of behaviors and acts, including nonverbal communication (Ekman & Friesen, 1981).

> These are acts which maintain the back-and-forth nature of speaking and listening between two or more interactants. They tell the speaker to continue, repeat, elaborate, hurry up, become more interesting, less salacious, give the other a chance to talk, etc.... The most common regulator is the head nod, the equivalent of the verbal mm-hmm; other regulators in-

clude eye contacts, slight movements forward, small postural shifts, eyebrow raises, and a whole host of other nonverbal acts. (Ekman & Friesen, 1981, p. 90)

While feedback has been studied in a speech-act context (Kraut & Higgins, 1984), the feedback-suffused protocol evidence of actual speakers has proved difficult or impossible to model using traditional speech acts (Cohen, 1984; Clark & Wilkes-Gibbs, 1986; Clark & Schaefer, 1989). This means that things like the coordination of turn-taking can not be modeled with speech acts because the turns frequently do not contain utterances which could be classified as speech acts. In studying referring expressions, Cohen (1984) and Clark Wilkes-Gibbs (1986) approached this problem by breaking down breaking down exchanges into segments characterized by their goals. I suggest that this approach can be extended to the problem of conversational control: each fragmentary utterance embodies an act that reflects its own proper purposes. For control in particular, the acts are meta-locutionary in the sense that they are illocutionary acts concerned with the conversation itself.

Meta-locutionary model

A computational model of meta-locutionary acts has been developed in order to test and explore this approach to controlling interaction. The model has three principal aspects: (1) a taxonomy of illocutionary meta-acts, (2) a modal logic representation for the conversants' understanding of the state of the conversation, and (3) situationally activated operators that relate the acts to the conversants' states. Early considerations of metadiscourse are reviewed by Beauvais (1989), who defined a speech-act model of metadiscourse for text as an idenfication of expositive illocutionary acts. In contrast, the model of conversational control presented in this paper extends the theory of Carbonell (1982) as a computational implementation.

Taxonomy of acts

Extending the taxonomic outline proposed by Bach and Harnish (1979) for communicative illocutionary acts, the model contains a forest of taxonomies, where each tree is corresponds to a control aspect of the conversation. Figures 2 shows the acts concerning turn-taking used in the simulation reported in this paper. Other kinds of acts used in the simulation were repair of models and information. Further taxonomies would include aspects such as attention and reference.

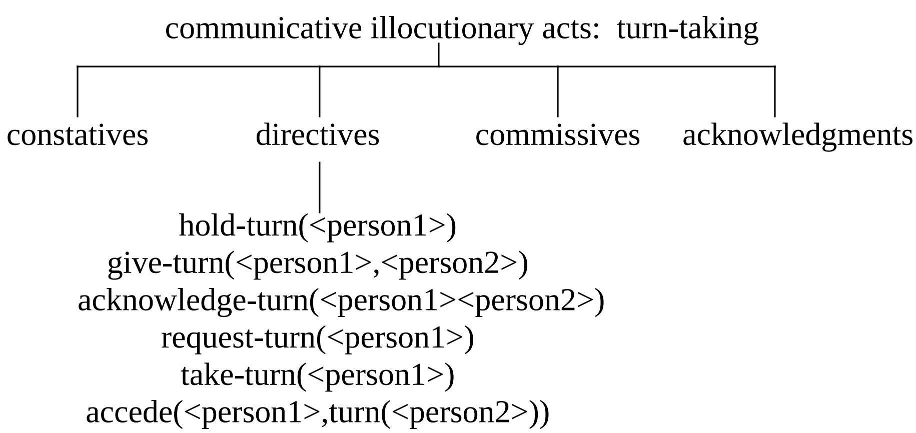

Figure 2. Taxonomy of meta-locutionary acts for turn-taking.

Modal representation for conversational state

The predicate representations for the conversants' understanding of their interaction must account for both domain knowledge and the conversational meta-knowledge discussed here. A modal representation with reasonably general expressiveness was chosen. Knowledge consists of beliefs and acts. Beliefs have the form believe(<agent>,<fact>,<truth-value>), where truth-value can be true, mutually-known-true, false, mutually-known-false, goal-true, goal-mutually-known-true, goal-false, and goal-mutually-known-false. The truth values reflect the mutual-knowledge view underlying the conversational model and the act-motivation view of speech-act theory. For example, believe(adam,(turn(barney),mutually-known-true) represents an agent's knowledge that Adam believes that he and Barney mutually know that it is Adam's turn. Knowledge of actions is represented in the form act(<agent>, <action>,<truth-value>). For example, the action act(adam,give-turn(barney),goal-true) indicates that Adam has as a goal to give the turn to Barney. This representation is discussed in more detail in (Novick, 1990).

Conversational operators

The third part of the model consists of operators that relate actions (including internal revision of beliefs) to the conversant's state. To carry on a conversation, the conversants must have both meta and domain operators. That is, some operators handle the control aspects of the conversation; it is with these that this paper is concerned. Other operators handle the pragmatic functions of dealing with the actual meaning of the interaction and its relation to the world.

For example, Figure 3 depicts the acknowledge-my-turn operator, as written with a Prolog-like syntax. The gist of this rule is that if one agent has given the turn to the other, then the recipient should make and acknowledging act. The "atts" part of the rule is for conflict resolution and documentation purposes. Other operators in the model correspond to cases of the other meta-acts presented in the taxonomy and to certain "internal" operations such as pruning redundant beliefs.

```
op(         % acknowledge_my_turn
 if([       agents(Me,[Other]),
            believe(Me,turn(Other),mutually_known_true),
            act(Other,give_turn(Me),true) ]),
 not([]),
 test([]),
 action([   act(Me,acknowledge_turn(turn(Me),mutually_known_true),true) ]),
 effects([  del(act(Other,give_turn(Me),true)),
            del(believe(Me,turn(Other),mutually_known_true)),
            add(believe(Me,turn(Me),mutually_known_true)) ]),
 atts([     level(turn),
            name(acknowledge_my_turn) ]) ).
```

Figure 3. Acknowledge-my-turn operator.

Simulation

As part of the empirical work in developing the model, protocols were obtained of conversations in a very simple domain. The conversation required for this research would have to contain minimal domain semantics, due to the difficulty of representing large amounts of domain knowledge and a need to avoid undetermined prior knowledge. At the same time, the conversation had to afford natural interaction control. Accordingly, protocols were taken of conversations where the conversants each had memorized a short partial sequence of random letters; together, the conversants could recall an entire, consistent sequence. Such conversations have been characterized as the "e. coli" of conversation analysis (Hamburger, 1989); they do, in fact, display virtually all of the traits of interaction described in the introduction to this paper.

Transcripts of the conversations, reported in Novick (1988), included notation of both verbal and physical states of the subjects. Possible illocutionary acts were noted, including both meta-acts and domain-acts.

A multi-agent, virtually parallel, rule-based system was developed to simulate the protocols using the meta-locutionary model. The details of the rule-based system are reported elsewhere (Novick, 1990). The simulation is in some ways similar to that of Power (1979), who modeled simple domain-level conversations between two robots. However, Powers found that his stack-based planning system for conversation would run into problems because of (1) incompleteness and (2) insufficient flexibility in adjusting to changes in context:

> Let us turn now to the second fault of the control stack as representation of the dialogue state: namely, insufficient explicitness. What this means is that the relations between elements of the dialogue state are not represented systematically.... The result is that the dialogue state can be interpreted just one way; it cannot be interpreted by several different procedures for several different purposes. The robots therefore cannot respond flexibly to unexpected turns in the conversation; an unexpected remark throws them completely. (Power, 1979, pp. 133-134).

In the current work, the meta-locutionary model, as implemented with situationally responsive operators, has largely overcome these problems. A set of domain-level operators was developed for the letter sequence domain. Results of the simulation suggest that the meta-locutionary model appears adequate for representation for control of conversational interaction with users. For the sequence domain, simple sequence conversations have been simulated successfully. For more extended conversations, difficulties in the simulation appear to be largely the consequence of the *domain* operators; the conversational control aspects remain robust.

Figure 4 shows a transcript of a simulated conversation between agents Adam and Barney. Adam has the sequence "_ I S" and Barney's sequence is "O _ S." Note that on each cycle, both agents' actions are taken simultaneously; the transcript simply has to show one agent's actions before the other.

Conclusion

Simulation of conversations in simple domains such as letter sequences suggests including meta-locutionary knowledge in user models should provide systems with more flexible and robust interaction. Mutual models of conversational knowledge can be seen as including both the domain knowledge that lets users "know what they're talking about" and the meta knowledge that lets users create their conversational "ensemble" interaction. This work is currently being extended to a more complex domain—air traffic control. Preliminary comparative analysis of air traffic control (ATC) with the letter sequence domain suggests that for ATC control may be simpler than for the face-to-face letter conversations. Further work should indicate whether this is a function of the domain, the modality, or both.

```
adam's acts for cycle 1:
  acknowledge_my_turn --
      act(adam,acknowledge_turn(turn(adam),mutually_known_true),true)
  goal_request_next_subsequence --
barney's acts for cycle 1:
  goal_confirm_next_subsequence --
adam's acts for cycle 2:
  do_request --
      act(adam,request(act(barney,assert(subsequence(1,[blank]),adam),
goal_true),barney),true)
  give_turn_3 --
      act(adam,give_turn(barney),mutually_known_true)
barney's acts for cycle 2:
  others_turn_acknowledged --
adam's acts for cycle 3:
  prune_redundant_beliefs --
barney's acts for cycle 3:
  recognize_my_turn_1 --
  goal_assert_next_subsequence_2 --
adam's acts for cycle 4:
barney's acts for cycle 4:
  do_assert --
      act(barney,assert(believe(barney,subsequence(1,[o]),true),adam),true)
adam's acts for cycle 5:
  assertion_received_1 --
  informed_of_next_subsequence_by_other --
barney's acts for cycle 5:
  asserted_next_subsequence --
adam's acts for cycle 6:
  goal_confirm_next_subsequence --
barney's acts for cycle 6:
  goal_request_next_subsequence --
adam's acts for cycle 7:
barney's acts for cycle 7:
  do_request --
      act(barney,request(act(adam,assert(subsequence(2,[blank]),barney),
                                  goal_true),adam),true)
  give_turn_3 --
      act(barney,give_turn(adam),mutually_known_true)
adam's acts for cycle 8:
  recognize_my_turn_1 --
  goal_assert_next_subsequence_2 --
barney's acts for cycle 8:
  prune_redundant_beliefs --
adam's acts for cycle 9:
  do_assert --
      act(adam,assert(believe(adam,subsequence(2,[i,s]),true),barney),true)
barney's acts for cycle 9:
adam's acts for cycle 10:
  asserted_last_subsequence --
barney's acts for cycle 10:
  assertion_received_1 --
  informed_of_last_subsequence_by_other --
adam's acts for cycle 11:
barney's acts for cycle 11:
```

Figure 4: Transcript of simulated conversation in letter sequence domain.

Acknowledgments

Fred Shaudys and Carolyn Sousa worked on refining the domain operators for the letter-sequence domain simulation. Karen Ward participated in the preliminary analysis of the air traffic control domain.

References

Bach, K., & Harnish, R. M. (1979). *Linguistic communication and speech acts*. Cambridge, MA: The MIT Press.

Beauvais, Paul J. (1989). A speech act theory of metadiscourse. *Written Communication*, **6**, 11-30.

Carbonell, J. G. (1982). *Meta-Language Utterances in Purposive Discourse*. Technical Report CMU-CS-82-125, Carnegie-Mellon University, Department of Computer Science.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10-63). New York: Cambridge University Press.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, **13**, 259-294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*. **22**, 1-39.

Cohen, P. R. (1984). The pragmatics of referring and the modality of communication. *Computational Linguistics*, **10**(2), 97-146.

Duncan, S., Jr. (1980). Describing face-to-face interaction. In W. Von Raffler-Engel (Ed.), *Aspects of nonverbal communication* (pp. 67-80). Lisse, Germany: Swets and Zeitlinger.

Ekman, P., & Friesen, W. V. (1981). The repetoire of nonverbal behavior. In A. Kendon (Ed.), *Nonverbal communication, interaction, and gesture* (pp. 57-105). The Hague: Mouton. (Reprinted from Semiotica, 1969, 1, 49-97.)

Kraut, R. E., Lewis, S. H., & Swezey, L. W. (1982). Listener responsiveness and the coordination of conversation. *Personality and Social Psychology*, **41**, 718-731.

Kraut, R. E., & Higgins, E. T. (1984). Communication and social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition*, **3**, (pp. 87-127). Hillsdale, NJ: Lawrence Erlbaum Associates

Novick, D. (1988). *Control of mixed-initiative discourse through meta-locutionary acts: a computational model*, Technical Report CIS-TR-88-18, Department of Computer and Information Science, University of Oregon, Eugene, Oregon, 1988.

Novick, D. (1990). Modeling belief and action in a multi-agent system, *Conference on AI, Simulation and Planning in High-Autonomy Systems*, Tucson, AZ, March, 1990

Power, R. (1979). The organization of purposeful dialogues. *Linguistics*, **17**, 107-152.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, **50**, 696-735.

Suchman, L. A. (1987). *Plans and situated actions*. Cambridge: Cambridge University Press.