DESIGN FOR DIALOGUE COMPREHENSION

William C. Mann
USC Information Sciences Institute
Marina del Rey, CA
April, 1979

This paper describes aspects of the design of a dialogue comprehension system, DCS, currently being implemented. It concentrates on a few design innovations rather than the description of the whole system. The three areas of innovation discussed are:

1. The relation of the DCS design to Speech Act theory and Dialogue Game theory,

2. Design assumptions about how to identify the "best" interpretation among several alternatives, and a method, called Preeminence Scheduling, for implementing those assumptions,

3. A new control structure, Hearsay-3, that extends the control structure of Hearsay-II and makes Preeminence Scheduling fairly straightforward.

I. Dialogue Games, Speech Acts and DCS -- Examination of actual human dialogue reveals structure extending over several turns and corresponding to particular issues that the participants raise and resolve. Our past work on dialogue has led to an account of this structure, Dialogue Game theory [Levin & Moore 1978; Moore, Levin & Mann 1977]. This theory claims that dialogues (and other language uses as well) are comprehensible only because the participants are making available to each other the knowledge of the goals they are pursuing at the moment. Patterns of these goals recur, representing language conventions; their theoretical representations are called Dialogue Games.

If a speaker employs a particular Dialogue Game, that fact must be recognized by the hearer if the speaker is to achieve the desired effect. In other words, Dialogue Game recognition is an essential part of dialogue comprehension. Invoking a game is an act, and terminating the ongoing use of a game is also an act.

Dialogue game theory has recently been extended [Mann 1979] in a way makes these game-related acts explicit Acts of Bidding a game, Accepting a bid, and Bidding termination are formally defined as speech acts, comparable to others in speech act theory. So, for example, in the dialogue fragment below,

C: "Mom, I'm hungry."

M: "Did you do a good job on your Geography homework?"

the first turn bids a game called the Permission Seeking game, and the second turn refuses that bid and bids the Information Seeking game.

DCS is designed to recognize people's use of dialogue games in transcripts. For each utterance, it builds a hierarchial structure representing how the utterance performs certain acts, the goals that the acts serve, and the goal structure that makes the combination of acts coherent. (The data structure holding this information is described below in the discussion of Hearsay-3.)

II. Preeminence Scheduling -- It seems inevitable that any system capable of forming the "correct" interpretation of most natural language usage will usually be able to find several other interpretations, given enough opportunity. It is also inevitable that choices be made, implicitly or explicitly, among interpretations. The choices will correspond to some internal notion of quality, also possibly implicit. The notion of quality may vary, but the necessity of making such choices does not rest on the particular notion of quality we use. Clearly, it is also important to avoid choosing a single interpretation when there are several nearly equally attractive ones.

What methods do we have for making such choices? Consider three approaches:

1. First-find: The first interpretation discovered which satisfies well-formedness is chosen. The effectiveness of first-find depends on having well-informed, selective processes at every choice point, and is only reasonable if one's expectations about what might be said are very good. Even then, this method will select incorrect interpretations.

2. Bounded search and ranked choice: Interpretations are generated by a bounded-effort search, each is assigned an individual quality score of some sort, and the best is chosen. While this will not miss good but unexpected interpretations missed by first-find, it is wrong in at least two ways: a) it selects an interpretation (and discards others) when the quality difference between interpretations is insignificant, and b) it expends unnecessary resources making absolute quality judgments where only relative judgments are needed. These defects suggest an improvement:

3. Preeminence selection: perform a bounded-effort search for interpretations, and then select as best the one (if any) having a certain threshold amount of demonstrable preferability over its competitors. The key to correct choice is determination that such a threshold difference in quality exists. DCS is designed to identify preeminent interpretations.

Consider the information content in the fact that the best two interpretations have a quality difference exceeding a fixed threshold. This fact is sufficient to choose an interpretation, and yet it carries less information than is carried in a set of quality scores for the same set of interpretations. Computational efficiencies are available because the work of creating the excess information can be avoided by proper design.

Given a tentative quality scoring of one's alternatives, several kinds of computations can be avoided. For the highest-ranked interpretation, it is pointless to perform computations whose only effect is to confirm or support the interpretation, (even though we expect that for correct interpretations the ways to show confirmation will be numerous), since these will only drive its score higher.

For interpretations with inferior ranks, it is likewise pointless to perform computations that refute them (although we expect that refutations of poor interpretations will be numerous), since these will only drive their scores lower. Neither of these is relevant to demonstrating preeminence.

Given effective controls, computation can concentrate on refuting good interpretations and supporting weak ones. (Of course, such computations will sometimes move a new interpretation into the role of highest-ranked. They may also destroy an apparent preeminence.) If the gap in quality rating between the highest ranked interpretation and the next one remains significant, then preeminence has been demonstrated.

Further efficiencies are possible provided that the maximum quality rating improvement from untried support computations can be predicted, since it is then possible to find cases for which the maximum support of a low-ranked interpretation would not eliminate an existing preeminence. Similar efficiencies can arise from predicting the maximum loss of quality available from untried refutations. This approach is being implemented in DCS.

III. Control Structure -- a new AI programming environment called Hearsay-3 is being implemented at ISI for use in development of several systems. It is an augmentation and major revision of some of the control and data structure ideas found in Hearsay-II [Lesser & Erman 1977], but it is independent of the speech-understanding task. Hearsay-3 retains interprocess communication by means of global "blackboards," and it represents its process knowledge in many specialized "knowledge source" (KS) processes, which nominate themselves at appropriate times by looking at the blackboard, and then are opportunistically scheduled for execution. Blackboards are divided into "levels" that typically contain distinct kinds of state knowledge, the distinctions being used as a gross filter on which future KS computations are considered.

Hearsay-3 retains the idea of a domain-knowledge blackboard (BB), and it adds a knowledge source scheduling blackboard (SBB) as well. Items on the SBB are opportunities to exercise particular scheduling specialists called Scheduling Knowledge Sources (SKS).

The SBB is an ideal data structure for implementing Preeminence scheduling. In DCS the SBB has four levels, called Refutation, Support, Evaluation and Ordinary-consequence. These correspond to a factoring of the domain KS into four groups according to their effects. Knowledge sources in each of these groups nominate themselves onto a different level of the SBB. The scheduling-knowledge sources (SKS) perform preeminence scheduling (when a suitable range of alternatives is available) by selecting available Refutation level opportunities for the highest-ranked interpretation and Support level opportunities for inferior ones. (The SBB and SKS features of Hearsay-3 are only two of its many innovations.)

The DCS BB has 6 levels, named Text, Word-senses, Syntax, Propositions, Speech-acts and Goals. Goals and goal structures, which are required in any successful analysis, only arise as explanations of speech acts. The KS used for deriving speech acts from utterances are separate from those deriving goals from speech acts. The hierarchic data structure representing an interpretation of an utterance consists of units at various levels on the Hearsay-3 blackboard.

USING DCS

These innovations and several others will be tested in DCS in attempts to comprehend human dialogue gathered from non-laboratory situations. (One of these is Apollo astronaut to ground communication.) Transcripts of actual interpersonal dialogues are particularly advantageous as study material, because they show the effects of ongoing communication and because they are free of the biases and narrow views inevitable in made-up examples.

ACKNOWLEDGMENTS

The work reported here was supported by NSF Grant MCS-76-07332.

REFERENCES

Lesser, V. R., and L. D. Erman, "A Retrospective View of the HEARSAY-II Architecture," *Fifth International Joint Conference on Artificial Intelligence*, Cambridge, MA, 1977.

Levin, J. A., and J. A. Moore, "Dialogue Games: Meta-communication Structures for Natural Language Interaction," Cognitive Science, 1,4, 1978.

Moore, J. A., J. A. Levin, and W. C. Mann, "A Goal-oriented Model of Human Dialogue," American Journal of Computational Linguistics, microfiche #67, 1977.

Mann, W. C., "Dialogue Games," in MODELS OF DIALOGUE, K. Hintikka, et al (eds.) North Holland Press, 1979.