SPEAKING WITH MANY TONGUES: SOME PROBLEMS IN MODELING SPEAKERS OF ACTUAL DISCOURSE

John H. Clippinger, Jr.
Teleos
Cambridge MA 02138

I. INTRODUCTION

> "It is as if our languages were confounded; when we want a thought, they bring us a word; when we ask for a word; they bring us a dash; and when we expect a dash, there comes a piece of bawdy."
> Lichtenberg

Any discussion of natural language processing models eventually makes reference to the extent to which a particular model or class of models accurately describes an aspect of human language processing. While there has been considerable work done in natural language comprehending systems and subsequent discussion of their appropriateness as descriptive models, there has not been a parallel effort in the area of discourse generation. Moreover, most work done in language comprehension and generation systems has been highly restricted to very specialized speech situations, and has not seen as its primary objective the modeling of a specific speaker or listener. While there are good methodological reasons for working with restricted examples and limited domains, an argument can be made that such a restricted and normative focus can result in distorted notions about the character and complexity of natural speech generation and comprehension.

In this paper I would like to break with some established practices in artificial intelligence and linguistics, and examine some of the properties of transcribed natural discourse in order to suggest the types of mechanisms and frameworks that might be most appropriate to a description of human discourse behavior. The paper consists of three parts: the first will begin with a brief analysis of a sample of a "therapeutic discourse" and then focus on some principal issues in modeling natural discourse; the following section will briefly describe the manner in which this particular discourse episode was modeled in the program ERMA, and the final section will discuss the limitations of my own approach and suggest others that may be more complete.

II. AN EXAMPLE OF NATURAL DISCOURSE

For all the work done in linguistics, it is only relatively recently that linguists -- especially sociolinguists -- have begun to look at actual language use. Like undergraduates in psychology experiments, there has been a tendency to use limited samples of speakers as data sources -- more often than not, linguists themselves. Although this trend is beginning to reverse itself, it nonetheless has resulted in some distortions about the use and character of language. For by solely focusing upon the grammaticality or well-formedness of an utterance, attention is often averted from the reasons a particular utterance was made, and why it took the particular form it did. Below is a fragment of a transcribed episode in therapeutic discourse which was used in constructing the ERMA model and which I believe, is rich in data about discourse behavior.

> You know, for some reason, I, huh, just thought about, uh, about the bill and payment again. That (pause 2 seconds) that, uh (pause 4 seconds) I was, uh, thinking that I -- of asking you whether it wouldn't be all right for (pause 2 seconds), you know, not to give me a bill. That is, uh-I would (hesitates) since I usually by -- well, I immediately thought of objections to this, but my idea was that I would simply count up the number of hours and give you a check at the end of the month.

Notice that the information content of this reques is quite limited; the patient wants to bill herself and pay by check at the end of the month. But this rather limited request is couched in a most extraordinary network of dubitatives -- "you know," "for some reason," "wouldn't be all right," "that is," "was thinking," etc. Similarly, the entire third sentence acts like a dubitative, and in effect was produced in anticipation of the therapist's reaction; obviously, the patient is worried about what her therapist might think. Notice too that the patient starts a sentence ("That is, uh-I would since I usually by-well,") and then initiaties another sentence. It is quite apparent that this particular speaker is monitoring most of what she says -- thinking about it, and then reacting to it. Moreover, she does not seem to know what she is finally going to say from the start, but modifies, qualifies, and corrects herself as she speaks. Hence she seems to be invoking a number of different and often contradictory criteria in her production of discourse. She doesn't so much generate discourse -- as that implies too neat and linear a process -- as she regulates it; she makes a statement, retracts it, qualifies it, and then restates it. In fact, as the remaining section of this discourse episode makes clear, she had only a limited understanding about why she made the request in the first place. It turns out that she initiated the request to reduce the formality of the therapeutic setting in order to facilitate "intercourse" between herself and the therapist. Not only would it appear that there are a number of different and conflicting control mechanisms operating during the course of discourse production, but also during its inception, when the notion of the request was first being considered.

What I think that this example demonstrates is the extent to which

cognitive and linguistic behaviors are incredibly intertwined. For not only does the cognitive component itself appear to be fragmented and at odds with itself, but each of these fragments -- or control mechanisms -- seem to have access to linguistic knowledge, which they in turn can use for their own purposes. Hence entire constituents can be formed and inserted en masse, ("you know," "for some reason",) or entire sentences generated by relatively independent mechanisms. Moreover, similar linguistic items seem to have different meanings according to how they are used, and therefore would require multiple representations. For example, the first sentence is a type of foregrounding remark and was constructed to acquaint the listener, the therapist in this case, with the reason for a change in topic. The word "thought" here refers to a prior mental action of the speaker and differs considerably from the "was thinking" of the second sentence, which in this case acts as a qualifier or dubitative. Any intelligent system must of course know the difference of uses and hence, meanings, between the two. Moreover, the derivation of the two similar words would differ; the first one would be derived from the concept for thinking and therefore come from the cognitive component proper, whereas the second has more of a thematic meaning and would be inserted as a means of qualifying a statement.

While it is quite possible to model or describe isolated instances of such examples, I am quite skeptical about the long term value of such models independent of a more general framework or theory. There is, I believe, a need for more global theories about what discourse is and does, for without such a framework, there is a tendency for classificatory schemes to become too closed and too specialized and hence incompatible with related work. This is particularly true if the process of discourse formation is a highly interdependent one, as I believe it is, and decisions made at one level interact with those made at another.

A more global theory of discourse would by necessity have to take into account the fact that discourse behavior is an intentional activity which attempts to affect some trade off between the speaker's goals, his constraints, his competency, and those of his audience. For it is within this setting that the more manifest aspects of discourse or speech are cast and it is within this setting that linguistic options, rules, and forms are organized. This is particularly true of thematic information which is especially variable and sensitive to the intentions and idiosyncracies of the speaker and his setting. Consequently, in order to understand why certain thematic forms were used, as against others, or why a particular form is appropriate in one context and not another, it is necessary to understand both the multiple intentions of the speaker and the constraints he or she felt were in effect. To some extent, to answer such questions entails a move from linguistics to psychology -- maybe analytic psychology at that. But independent of such a move I am at a loss to see how discourse behavior can be adequately described at the linguistic level alone.

III. THE ERMA MODEL OF DISCOURSE BEHAVIOR

The ERMA (err-umm-ah) (Clippinger, 1974; Brown, 1974), program was written to simulate a speaker of a therapeutic discourse from beginning to end: the motivation of the discourse, its censoring, reformulation, expression, and introspection. The intent was to have a program which replicated the thought and discourse processes of the modeled speaker -- including her hesitations and mistakes. The program was initially written in CONNIVER (McDermott Sussman, 1973) and makes extensive use of two of its key features; methods and contexts.

Five major stages of the discourse formation and expression process were identified and represented as CONNIVER contexts: Calvin, Machiavelli, Cicero, Freud, and Realization. Each had their own programs and datums, and accordingly, often their own opinions about what should be said and how it should be said. The discourse stream, so to speak, has its source in a special program which initiates topic concepts for action, and then flows through each of these contexts -- often back and forth before achieving its final expression. When a concept (more will be said about what a concept is later) is placed within the Calvin context, programs there examine that concept to determine its acceptability for expression. The concept can be censored right off or it can be passed on to the Machiavelli context with suggestions as to how it should be modified. For example, in the case of the discourse sample previously discussed, the motivating concept expressed a desire for intercourse with the therapist. This request is found to be unacceptable, and the reason given is the fact that the therapy situation is a formal one governed by the exchange of payments. Another program in Calvin then tries to see what would happen if the exchange relationship were "negated" or eliminated. It sends this new concept to the Machiavelli context, who as a specialist in such matters, tries to figure out how this would be accomplished. Various programs there consider different possibilities, one of which is to ask the therapist not to give the patient a bill. This suggestion in turn is sent to the Cicero context to assess its possible impact upon the therapist. A program there says that it would shock the therapist and sends back the concept for further consideration. Once back in Machiavelli again, programs there, who specialize in reducing negative or shocking impacts of statements, make their suggestions as to how the request can be made, with a minimum of offense. It is at this point in the program that cognitive and thematic considerations become mixed. For although the program is still working with concepts, which may or may not have acceptable lexical realization, it is making suggestions as to how the concepts should be marked for expression and how certain

qualifying constituents can be inserted to modify the listener's interpretation. These suggestions are then again sent to the Cicero context, where barring no further objections, they are made into state concepts and sent onto the Realization context for lexical realization and ordering. However, even after lexical realization is completed, the possibility exists for further interruption and reformulation, as Calvin has one final look at what is to be said. He can either stop an expression altogether, for example, where the patient starts to say "I shouldn't be given a bill" but stops herself and instead says "of asking you whether it wouldn't be...", or he can make last minute suggestions, which in turn are translated by special programs into interjections such as, "you know," "huh," "I mean."

In order to produce even this limited discourse the program had to be able to distinguish between several levels of cognition: thinking about how to achieve a particular goal (irrespective of whether it would be realized into discourse); thinking about what to say (a state concept) and thinking about how to say it (lexical realization). At each of these levels different points of view were involved; Calvin regularly objected to what Cicero found acceptable, and Machiavelli, though apolitical, often found himself caught in between. If one adds in the Freud context, used for introspection, then the clash of opinions can become all the more strident. While it is doubtful that all forms of discourse behavior involve such a multitude of personalities and agents, it is apparent, I believe, that discourse can be produced from a number of different sources and that higher level cognitive decisions effect lower level linguistic decisions and vice versa.

Since the discourse production process was found to be so diverse and interactive, it was absolutely necessary to have an extremely flexible means for representing concepts. The basic notion is similar to that of frames (Minsky, 1974; Winograd, 1974) where a "generic concept" is used to represent a general or stereotypic meaning and a "token concept" is used to represent an instantiation of the generic concept. House, for example, would be treated as the generic concept, and the big red house would be the token concept where "big" and "red" would be concepts on the property list. Information about what the concept was would be carried on one set of indicators such as CHARACTERISTIC, and how it was to be used by another, DENOTER, PROCESS, DE-EMPHASIZE, STRESS, etc. Similarly, information about concepts' relationship to other concepts would also be represented by indicators and their property lists: RELATED-CONCEPT, MOTIVATION, CAUSE, EFFECT, UNACCEPTABLE, etc. In all, about thirty different types of indicators were used.

Token concepts could be combined together to form "conceptual clauses," where a process or "relator" concept occupied the primary position and all other concepts filled the slot of arguments specifying a role with respect to the relator concept. These conceptual clauses could become arbitrarily complex where each argument could potentially contain other embedded conceptual clauses. Consequently, it was quite possible for the program to contain a thought which could not be expressed with the same meaning it had internally.

All processing of concepts is performed by a modified form of CONNIVER methods called multiple body interrupts developed by Richard Brown (Brown, 1974). They are fired when a concept's particular pattern matches their invocation pattern within a given context. Therefore control within the program is transferred through the sending and receiving of concepts and the addition and removal of property lists and indicators. For the most part, the intelligence of these programs is very specialized, as they only worked with token concepts for relatively specific purposes. Certainly if the program were to be expanded and generalized, these programs would also have to be generalized, but how much I am uncertain. For example, it might be tempting to have one program perform all foregrounding. However, I think that this would be disadvantageous in the long run, as the type of foregrounding required in discourse seems to vary according to the context in which the statement is being prepared, and therefore should be performed locally.

The organization of the grammar, I believe, also demonstrates the extent to which cognitive and linguistic factors are interdependent. Initially, I had planned to use Halliday's systemic grammar (Halliday, 1968), as I found his functional representation of grammar, and especially his work on thematic structure to be consistent with my own thinking about discourse structure. However, as I began to adapt it to the model I found that his level of description was inappropriate to what I was attempting to do. The first problem was that it was not conceptually based, and therefore too tightly wedded to a rather closed description of linguistic structures at the clausal level. The second problem was that his descriptions of thematic structure were not strictly functional, and therefore rather than supplying information about how a particular construction could be used to achieve a particular effect, they instead told what it was. Information about whether a particular concept is given or new within the ERMA program, for example, is not given a priori but supplied through feedback by audience sensitive programs that comment on what the audience knows or can be expected to know. Consequently, if the program is acting half way intelligently and tracking on what it is saying, it will take care of matters such as topic and comment implicitly. It is my belief that notions like theme and rheme, topic and comment, and the alike are only known upon reflection about what has been said, and essentially are tangential to the interests of the speaker while speaking. The grammar I ended up designing is conceptually based, and

while borrowing from Fillmore's notion of cases (Fillmore, 1968), employes a classification scheme which is ordered according to the effects various linguistic options achieve rather than simply what they are.

Conjectures and Conclusions

While the study of discourse structure and cognition is now being actively pursued by linguists (Hymes and Gumperz, 1972; Labov, 1975; Lakoff, 1971; Halliday, 1973), philosophers (Ricoeur, 1970, Searle, 1971), and computational linguists (MacDonald, 1975; Hayes, 1973; Schank, 1974; Philips, 1973, Schmidt, 1973), there have not to my knowledge been any models developed which attempt to model the discourse behavior of a specific speaker. As a consequence, it is difficult to place the ERMA model or theory within any given paradigm of research on discourse, as the frameworks, objectives, and methods within this field are as nearly varied as the people working in it. Moreover, it is my belief that computational linguistics -- not to mention discourse simulation -- has yet to find its paradigm, for there remains a welter of data yet to be organized or explained, and no overriding consensus as to how it should be done.

Therefore seen from this light, a discussion of the limitations and the goals of the ERMA model assumes a somewhat different cast, as I believe that the local and more technical decisions made in developing the model are of less relevance to an integrated understanding of human discourse than are some basic theoretical decisions which were made. For at present there is no basic methodological consensus on how discourse should be described -- or even what it is, for that matter, and hence any subsequent discussion of mechanisms alone would be, I believe, ungrounded. Therefore, I first want to make clear my own theoretical inclinations and then discuss some of the approaches of others to see how disourse behavior can be more comprehensively represented. First off, I regard the discourse process as a regulatory one, whereby the speaker attempts to achieve some effect through his communication of the medium of language. In speaking he has intentions as well as counter-intentions which he attempts to realize under the constraints of his own intellectual and linguistic competency as well as those of his listener. He can listen to what he says, and he can modify what he says at virtually any point in the discourse stream. He makes his linguistic selections according to the communicative effect he feels they will have; hence he is less concerned with the grammaticality or well-formedness of his utterance than with their effectiveness; in fact he might find being ungrammatical an effective means of expression. Moreover, he can be inventive as he speaks and listens, devising more abbreivated and specialized expressions according to context and intent.

In developing the ERMA model I attempted to embody as many of these notions of discourse behavior as possible, but I found that as I got further into working out a representation of the processes involved in producing the sample text, that I needed more empirical evidence for my modeling decisions. For example, why does the speaker hesitate at a particular point and not another, or why does she use "I mean," as against "You know," or how do I know that her request for a change in the billing arrangement is really a displaced request for intercourse? While I did examine other transcripts, roughly two hundred of them, to detect patterns, I lacked a sufficiently developed framework to really organize and focus my search. This lack of both empirical evidence and a more comprehensive and systematic understanding of discourse behavior in general led me to represent some of the discourse mechanisms and processes in a somewhat ad hoc fashion. For example, for the purposes of a general theory, it would have been desirable to segregate in more detail the different possible stages of discourse formation and production and be able to justify these distinctions on sound theoretical or empirical grounds. Similarly, I would have liked to have had some basis for classifying the types of roles indicators play, and hence provide a more general understanding of their part in shaping discourse. But as it were, there were so many sizable gaps in my knowledge about natural discourse that it was far too easy to make any number of simplifying, if not erroneous, assumptions about how discourse was formed and produced without any immediate penalty.

Consequently I would like to make an argument for more empirical research into the character of natural-unedited discourse, preferably those forms of discourse that exhibit mistakes and errors. I think that such inquiries in turn will provide a basis for designing and selecting the representational structures and procedures appropriate to model natural discourse. Certainly this cannot be done independent of some more general computational theory, but it is critical that this theory be not too specified or closed to preclude a full view of discourse data.

Finally, I would like to briefly comment on two recent approaches in artificial intelligence which might begin to fill in the gaps in my own work. One of the hazards of research in this area is that by the time you complete your work someone else has already come up with an improvement on your own ideas. The notion of frames (Minsky, 1974; Winograd, 1975) is a case in point. Although ERMA uses a knowledge representation similar to that of frames, her cognitive skills are too specialized and too localized; she is quite incapable of making inferential loops. Moreover, the ERMA model has no systematic overall organization to her belief system, and hence, it is difficult at times to visualize from a distance what is taking place. There is also the question as to whether the production of discourse is best organized according to relatively set contexts, such as Machiavelli and Calvin, or whether the concepts themselves should determine their

own formation and production contexts. I can see how something like a frame's representation might be able to resolve some of these problems because of its means for integrating general and specific knowledge. But exactly how I am uncertain. Perhaps, instead of having the program think in terms of concepts, it would think in terms of frame systems, which in turn would carry their own opinions as to how the discourse should be expressed. Similarly, these frame systems would be a part of larger frame systems, which would ovesee the overall and more general aspects of discourse regulation.

I would also like to adapt some key notions of Sussman's model of skill acquisition (Sussman, 1973) to a discourse model, as I feel that the process of producing a discourse and acquiring a skill -- albeit manipulating blocks, are intrinsically similar. Morever, his notions of representing actions in terms of their effects as well as his treatment of learning as a feedback process involving debugging and patching procedures, is similar in kind to my own notion of discourse production as a feedback regulated process. However, the added advantage of his approach is that the knowledge that his model has or acquires is a function of its own experience. This is substantially different from ERMA, where her knowledge is not learned to any substantial degree but programmed in from the start. This I feel to be a basic limitation, as it is my belief that at least part of an understanding of what discourse behavior is entails understanding how it was derived; especially, if one is concerned with the more subtle psychological factors that underlie therapeutic dicourse. For example, the manner in which the discourse formation contexts and procedures are organized reflect the learning and the experience of the speaker over time. Consequently, if we are ever going to adequately explain the organization of a discourse formation process of a real speaker, we are going to have to be able to identify those mechanisms which generated the discourse procedures and contexts in the first place. This is not just because it would be nice to have a means of describing the evolution of discourse styles and conventions over time, but because a single speaker will in the course of speaking use such deeper level mechanisms to both invent and to interpret new discourse.

In conclusion, I would like to encourage the developement of more comprehensive computational theories of discourse that link diachronic descriptions and concerns with synchronic ones. If this is done with sufficient empirical content, I feel it is quite possible to create good descriptive and perhaps even predictive models of human discourse behavior in the near future.

REFERENCES

Brown, R., The Use of Multiple Body Interrupts in Discourse Generation, Unpublished Undergraduate Thesis, MIT, 1974.

Clippinger, J., A Discourse Speaking Program as a Preliminary Theory of Discourse Behavior and a Limited Theory of Psychoanalytic Discourse, unpublished dissertation, University of Pennsylvania, 1974.

Fillmore, C., "The Case for Case" in Bach, E., Harms, R., eds., Universals in Linguistic Theory, Holt, Rinehart, and Winston, 1968.

Halliday, M.A.K., "Notes on Transitivity and Theme in English," Journal of Linguistics, 3, 1967, p. 37-81, 119-244, 179-216.

----- "The functional basis of language," in Bernstein, B., ed., Class, Codes and Control, Vol. II, Routledge, Kegan, and Paul, London, p. 343-365, 1973.

Hays, D., "Language and Interpersonal Relationships," Daedalus, Summer, 1973, p. 203-217.

-----Textual Organization, MS., 1973.

Hymes, D., Gumperz, J., eds, Directions in Sociolinguistics: The Ethnography of Communication, Holt, Rinehart and Winston, 1972.

Labov, W., Therapeutic Discourse, forthcoming.

Lakoff, G., "Conversational Postulates", Papers from the Seventh Regional Meeting of the Chicago Linguistic Society, Chicago Linguistic Society, 1971.

MacDonald, D., Preliminary Report on a Program for Generating Natural Languge, MIT AI Laboratory, 1975.

McDermott, D., Sussman, G., Conniver Reference Manual, MIT AI Laboratory, 1973.

Minsky, M., A Framework for Representing Knowledge, MIT AI Laboraaory, AI Memo No. 306, 1974.

Philips, B., "Topic Analysis," Proceedings of the 1973 International Conference on Computational Linguistics, 1973.

Ricoeur, P., "The Model of the Text: Meaningful Action as a Text," in Social Research, Vol. 38/Number 3, Autumn, 1971.

Schank, R., Understanding Paragraphs, Technical Report 5, Instituto per gli Studi Semantici e Cognitive, Castagnola, Switzerland, 1974.

Searle, J., ed., The Philosophy of Language, Oxford University Press, 1971.

Schmidt, C.F., D'Addamio, J., A Model of the Common Sense Theory of Intention, MS., 1973.

Sussman, G., A Computational Model of Skill Acquisition, MIT AI, TR-297, 1973.

Winograd, T., Frame Representations and the Declarative/Procedural Controversy, in Bobrow and Collins (eds) Representation and Understanding, Academic Press, 1975.