

DEPENDENCIES OF DISCOURSE STRUCTURE ON THE MODALITY  
OF COMMUNICATION: TELEPHONE vs. TELETYPE

Philip R. Cohen  
Dept. of Computer Science  
Oregon State University  
Corvallis, OR 97331

Scott Fertig  
Bolt, Beranek and Newman, Inc.  
Cambridge, MA 02239

Kathy Starr  
Bolt, Beranek and Newman, Inc.  
Cambridge, MA 02239

ABSTRACT

A desirable long-range goal in building future speech understanding systems would be to accept the kind of language people spontaneously produce. We show that people do not speak to one another in the same way they converse in typewritten language. Spoken language is finer-grained and more indirect. The differences are striking and pervasive. Current techniques for engaging in typewritten dialogue will need to be extended to accomodate the structure of spoken language.

I. INTRODUCTION

If a machine could listen, how would we talk to it? This question will be hard to answer definitively until a good mechanical listener is developed. As a next best approximation, this paper presents results of an exploration of how people talk to one another in a domain for which keyboard-based natural language dialogue systems would be desirable, and have already been built (Robinson et al., 1980; Winograd, 1972).

Our observations are based on transcripts of person-to-person telephone-mediated and teletype-mediated dialogues. In these transcripts, one specific kind of communicative act dominates spoken task-related discourse, but is nearly absent from keyboard discourse. Importantly, when this act is performed vocally it is never performed directly. Since most of the utterances in these simple dialogues do not signal the speaker's intent, techniques for inferring intent will be crucial for engaging in spoken task-related discourse. The paper suggests how a plan-based theory of communication (Cohen and Perrault, 1979; Perrault and Allen, 1980) can uncover the intentions underlying the use of various forms.

---

This research was supported by the National Institute of Education under contract US-NIE-C-400-76-0116 to the Center for the Study of Reading of the University of Illinois and Bolt, Beranek and Newman, Inc.

II. THE STUDY

Motivated by Rubin's (1980) taxonomy of language experiences and influenced by Chapanis et al.'s (1972, 1977) and Grosz' (1977) communication mode and task-oriented dialogue studies, we conducted an exploratory study to investigate how the structure of instruction-giving discourse depends on the communication situation in which it takes place. Twenty-five subjects ("experts") each instructed a randomly chosen "apprentice" in assembling a toy water pump. All subjects were paid volunteer students from the University of Illinois. Five "dialogues" took place in each of the following modalities: face-to-face, via telephone, teletype ("linked" CRT's), (non-interactive) audiotape, and (non-interactive) written. In all modes, the apprentices were videotaped as they followed the experts' instructions. Telephone and Teletype dialogues were analyzed first since results would have implications for the design of speech understanding and production systems.

Each expert participated in the experiment on two consecutive days, the first for training and the second for instructing an apprentice. Subjects playing the expert role were trained by: following a set of assembly directions consisting entirely of imperatives, assembling the pump as often as desired, and then instructing a research assistant. This practice session took place face-to-face. Experts knew the research assistant already knew how to assemble the pump. Experts were given an initial statement of the purpose of the experiment, which indicated that communication would take place in one of a number of different modes, but were not informed of which modality they would communicate in until the next day.

In both modes, experts and apprentices were located in different rooms. Experts had a set of pump parts that, they were told, were not to be assembled but could be manipulated. In Telephone mode, experts communicated via a standard telephone and apprentices communicated through a speaker-phone, which did not need to be held and which allowed simultaneous two-way communication. Distortion of the expert's voice was apparent, but not measured.

Subjects in "Teletype" (TTY) mode typed their communication on Elite Datamedia 1500 CRT

terminals connected by the Telenet computer network to a computer at Bolt, Beranek and Newman, Inc. The terminals were "linked" so that whatever was typed on one would appear on the other. Simultaneous typing was possible and did occur. Subjects were informed that their typing would not appear simultaneously on either terminal. Response times averaged 1 to 2 seconds, with occasionally longer delays due to system load.

#### A. Sample Dialogue Fragments

The following are representative fragments of Telephone and Teletype discourse.

##### A Telephone Fragment

S: "OK. Take that. Now there's a thing called a plunger. It has a red handle on it, a green bottom, and it's got a blue lid."

J: OK

.

.

.

S: OK now, the small blue cap we talked about before?

J: Yeah

S: Put that over the hole on the side of that tube —

J: Yeah

S: -- that is nearest to the top, or nearest to the red handle.

J: OK

S: You got that on the hole?

J: yeah

S: Ok. now. now, the smallest of the red pieces?

J: OK"

##### A Teletype Dialogue Fragment

B: "fit the blue cap over the tube end

N: done

B: put the little black ring into the large blue cap with the hiole in it...

N: ok

B: put the pink valve on the two pegs in that blue cap...

N: ok"

Communication in Telephone mode has a distinct pattern of "find the x" "put it into/onto/over the y", in which reference and predication are addressed in different steps. To relate these steps, more reliance is placed on strategies for signalling dialogue coherence, such as the use of pronouns. Teletype communication involves primarily the use of imperatives such as "put the x into/onto/around the y". Typically, the first time each object (X) is mentioned in a TTY discourse is within a request for a physical action.

#### B. A Methodology for Discourse Analysis

This research aims to develop an adequate method for conducting discourse analysis that will be useful to the computational linguist. The method used here integrates psychological, linguistic, and formal approaches in order to characterize language use. Psychological methods are needed in setting up protocols that do not bias the interesting variables. Linguistic methods are needed for developing a scheme for describing the progress of a discourse. Finally, formal methods are essential for stating theories of utterance interpretation in context.

To be more specific, we are ultimately interested in similarities and differences in utterance processing across modes. Utterance processing clearly depends on utterance form and the speaker's intent. The utterances in the transcripts are therefore categorized by the intentions they are used to achieve. Both utterances and categorizations become data for cross-modal measures as well as for formal methods. Once intentions differing across modes are isolated, our strategy is to then examine the utterance forms used to achieve those intentions. Thus, utterance forms are not compared directly across modes; only utterances used to achieve the same goals are compared, and it is those goals that are expected to vary across modes. With form and function identified, one can then proceed to discuss how utterance processing may differ from one mode to another.

Our plan-based theory of speech acts will be used to explain how an utterance's intent coding can be derived from the utterance's form and the prior interaction. A computational model of intent recognition in dialogue (Allen, 1979; Cohen, 1979; Sidner et al., 1981) can then be used to mimic the theory's assignment of intent. Thus, the theory of speech act interpretation will describe language use in a fashion analogous to the way that a generative grammar describes how a particular deep structure can underlie a given surface structure.

#### C. Coding the Transcripts

The first stage of discourse analysis involved the coding of the communicator's intent in making various utterances. Since attributions of intent are hard to make reliably, care was taken to avoid biasing the results. Following the experiences of Sinclair and Coulthard (1975), Dore et al. (1978) and Mann et al. (1975), a coding

scheme was developed and two people trained in its use. The coders relied both on written transcripts and on videotapes of the apprentices' assembly.

The scheme, which was tested and revised on pilot data until reliability was attained, included a set of approximately 20 "speech act" categories that were used to label intent, and a set of "operators" and propositions that were used to describe the assembly task, as in (Sacerdoti, 1975). The operators and propositions often served as the propositional content of the communicative acts. In addition to the domain actions, pilot data led us to include an action of "physically identifying the referent of a description" as part of the scheme (Cohen, 1981). This action will be seen to be requested explicitly by Telephone experts, but not by experts in Teletype mode.

Of course, a coding scheme must not only capture the domain of discourse, it must be tailored to the nature of discourse per se. Many theorists have observed that a speaker can use a number of utterances to achieve a goal, and can use one utterance to achieve a number of goals. Correspondingly, the coders could consider utterances as jointly achieving one intention (by "bracketing" them), could place an utterance in multiple categories, and could attribute more than one intention to the same utterance or utterance part.

It was discovered that the physical layout of a transcript, particularly the location of line breaks, affected which utterances were coded. To ensure uniformity, each coder first divided each transcript into utterances that he or she would code. These joint "bracketings" were compared by a third party to yield a base set of codable (sic) utterance parts. The coders could later bracket utterances differently if necessary.

The first attempt to code the transcripts was overly ambitious — coders could not keep 20 categories and their definitions in mind, even with a written coding manual for reference. Our scheme was then scaled back — only utterances fitting the following categories were considered:

Requests-for-assembly-actions (RAACT)  
(e.g., "put that on the hole".)

Requests-for-orientation-actions (RORT)  
(e.g., "the other way around", "the top is the bottom".)

Requests-to-pick-up (RPUP)  
(e.g., "take the blue base".)

Requests-for-identification (RID)  
(e.g., "there is a little yellow piece of rubber".)

Requests-for-other (ROTH)  
(e.g., requests for repetition, requests to stop, etc.)

Inform-completion(action)  
(e.g., "OK", "yeah", "got it".)

Label  
(e.g., "that's a plunger")

Interrater reliabilities for each category (within each mode), measured as the number of agreements X 2 divided by the number of times that category was coded, were high (above 90%). Since each disagreement counted twice (against both categories that were coded), agreements also counted twice.

#### D. Analysis 1: Frequency of Request types

Since most of each dialogue consisted of the making of requests, the first analysis examined the frequency of the various kinds of requests in the corpus of five transcripts for each modality. Table I displays the findings.

TABLE I

#### Distribution of Requests

Type	Telephone		Teletype	
	Number	Percent	Number	Percent
RAACT	73	25%	69	51%
RORT	26	9%	11	8%
ROTH	43	15%	18	13%
RPUP	45	16%	23	17%
RID	101	35%	13	10%
Total:	288		134	

This table supports Chapanis et al.'s (1972, 1977) finding that voice modes were about "twice as wordy" as non-voice modes. Here, there are approximately twice as many requests in Telephone mode as Teletype. Chapanis et al. examined how linguistic behavior differed across modes in terms of measures of sentence length, message length, number of words, sentences, messages, etc.

In contrast, the present study provides evidence of how these modes differ in utterance function. Identification requests are much more frequent in Telephone dialogues than in Teletype conversations. In fact, they constitute the largest category of requests — fully 35%. Since utterances in the RORT, RPUP, and ROTH categories will often be issued to clarify or follow up on a previous request, it is not surprising they would increase in number (though not percentage) with the increase in RID usage. Furthermore, it is sensible that there are about the same number of requests for assembly actions (and hence half the percentage) in each mode since the same "assembly work" is accomplished. Therefore, identification requests seem to be the primary request differentiating the two modalities.

#### E. Analysis 2: First time identifications

Frequency data are important for computational linguistics because they indicate the kinds of utterances a system may have to

interpret most often. However, frequency data include mistakes, dialogue repairs, and repetition. Perhaps identification requests occur primarily after referential miscommunication (as occurs for teletype dialogues (Cohen, 1981)). One might then argue that people would speak more carefully to machines and thus would not need to use identification requests frequently. Alternatively, the use of such requests as a step in a Telephone speaker's plan may truly be a strategy of engaging in spoken task-related discourse that is not found in TTY discourse.

To explore when identification requests were used, a second analysis of the utterance codings was undertaken that was limited to "first time" identifications. Each time a novice (rightly or wrongly) first identified a piece, the communicative act that caused him/her to do so was indicated. However, a coding was counted only if that speech act was not jointly present with another prior to the novice's part identification attempt. Table II indicates the results for each subject in Telephone and Teletype modes.

TABLE II

Speech Acts just preceding novices' attempts to identify 12 pieces.

Telephone				Teletype			
SUBJ	RID	RPUP	RAACT	RID	RPUP	RAACT	
1	9	2	1	1	2	9	
2	1	10	1	0	2	9	
3	11	1	0	1	2	9	
4	9	1	0	0	6	3	
5	10	0	0	2	6	4	

Subjects were classified as habitual users of a communicative act if, out of 12 pieces, the subject "introduced" at least 9 of the pieces with that act. In Telephone mode, four of five experts were habitual users of identification requests to get the apprentice to find a piece. In Teletype mode, no experts were habitual users of that act.

To show a "modality effect" in the use of the identification request strategy, the number of habitual users of RID in each mode were subjected to the Fischer's exact probability test (hypergeometric). Even with 5 subjects per mode, the differences across modes are significant ( $p = 0.023$ ), indicating that Telephone conversation per se differs from Teletype conversation in the ways in which a speaker will make first reference to an object.

F. Analysis 3: Utterance forms

Thus far, explicit identification requests have been shown to be pervasive in Telephone mode and to constitute a frequently used strategy. One might expect that, in analogous circumstances, a machine might be confronted with many of these acts. Computational linguistics research then must discover means by which a machine can determine the appropriate response as a function, in part, of the form of the utterance. To see

just which forms are used for our task, utterances classified as requests-for-identification were tabulated. Table III presents classes of these utterance, along with an example of each class. The utterance forms are divided into four major groups, to be explained below. One class of utterances comprising 7% of identification requests, called "supplemental NP" (e.g., "Put that on the opening in the other large tube. with the round top"), was unreliably coded and not considered for the analyses below. Category labels followed by "(?)" indicate that the utterances comprising those categories might also have been issued with rising intonation.

TABLE III  
Kinds of Requests to Identify in Telephone Mode

Group	CATEGORY [example]	Per Cent of RID's
-------	--------------------	-------------------

A. ACTION-BASED

- 1. THERE'S A NP(?) 28%  
["there's a black o-ring(?)"]
- 2. INFORM(IF ACT THEN EFFECT) 4%  
["If you look at the bottom you will see a projection"]
- 3. QUESTION(EFFECT) 4%  
["Do you see three small red pieces?"]
- 4. INFORM(EFFECT) 3%  
["you will see two blue tubes"]

B. FRAGMENTS

- 1. NP AND PP FRAGMENTS (?) 9%  
["the smallest of the red pieces?"]
- 2. PREPOSED OR INTERIOR PP (?) 6%  
["In the green thing at the bottom <pause> there is a hole"]  
["Put that on the hole on the side of that tube...that is nearest the top"]

C. INFORM(PROPPOSITION) → REQUEST(CONFIRM)

- 1. OBJ HAS PART 18%  
["It's got a peg in it"]
- 2. LISTENER HAS OBJ 5%  
["Now you have two devices that are clear plastic"]
- 3. DESCRIPTION1 = DESCRIPTION2 8%  
["The other one is a bubbled piece with a blue base on it with one spout"]

---

#### D. NEARLY DIRECT REQUESTS

["Look on the desk"]	2%
["The next thing your gonna look for is..."]	1%

---

Notice that in Telephone mode identification requests are never performed directly. No speaker used the paradigmatic direct forms, e.g. "Find the rubber ring shaped like an O", which occurred frequently in the written modality. However, the use of indirection is selective — Telephone experts frequently use direct imperatives to perform assembly requests. Only the identification-request seems to be affected by modality.

### III. INTERPRETING INDIRECT REQUESTS FOR REFERENT IDENTIFICATION

Many of the utterance forms can be analyzed as requests for identification once an act for physically searching for the referent of a description has been posited (Cohen, 1981). Assume that the action IDENTIFY-REF (AGT, DESCRIPTION) has as precondition "there exists an object O perceptually accessible to agt such that O is the (semantic) reference of DESCRIPTION." The result of the action might be labelled by (IDENTIFIED-REF AGT DESCRIPTION). Finally, the means for performing the act will be some procedural combination of sensory actions (e.g., looking) and counting. The exact combination will depend on the description used. The utterances in Group A can then be analyzed as requests for IDENTIFY-REFERENT using Perrault and Allen's (1980) method of applying plan recognition to the definition of communicative acts.

#### A. Action-based Utterances

Case 1 ("There is a NP") can be interpreted as a request that the hearer IDENTIFY-REFERENT of NP by reasoning that a speaker's informing a hearer that a precondition to an action is true can cause the hearer to believe the speaker wants that action to be performed. All utterances that communicate the speaker's desire that the hearer do some action are labelled as requests.

Using only rules about action, Perrault and Allen's method can also explain why Cases 2, 3, and 4 all convey requests for referent identification. Case 2 is handled by an inference saying that if a speaker communicates that an act will yield some desired effect, then one can infer the speaker wants that act performed to achieve that effect. Case 3 is an example of questioning a desired effect of an act (e.g., "Is the garbage out?") to convey that the act itself is desired. Case 4 is similar to Case 2, except the relationship between the desired effect and some action yielding that effect is presumed.

In all these cases, ACT = LOOK-AT, and EFFECT = "HEARER SEE X". Since LOOK-AT is part of the "body" (Allen, 1979) of IDENTIFY-REFERENT, Allen's "body-action" inference will make the necessary connection, by inferring that the speaker wanted the hearer to LOOK-AT something as part of his IDENTIFY-REFERENT act.

#### B. Fragments

Group B utterances constitute the class of fragments classified as requests for identification. Notice that "fragment" is not a simple syntactic classification. In Case 2, the speaker paralinguistically "calls for" a hearer response in the course of some linguistically complete utterance. Such examples of parallel achievement of communicative actions cannot be accounted for by any linguistic theory or computational linguistic mechanism of which we are aware. These cases have been included here since we believe the theory should be extended to handle them by reasoning about parallel actions. A potential source of inspiration for such a theory would be research on reasoning about concurrent programs.

Case 1 includes NP fragments, usually with rising intonation. The action to be performed is not explicitly stated, but must be supplied on the basis of shared knowledge about the discourse situation — who can do what, who can see what, what each participant thinks the other believes, what is expected, etc. Such knowledge will be needed to differentiate the intentions behind a traveller's saying "the 3:15 train to Montreal?" to an information booth clerk (who is not intended to turn around and find the train), from those behind the uttering of "the smallest of the red pieces?", where the hearer is expected to physically identify the piece.

According to the theory, the speaker's intentions conveyed by the elliptical question include 1) the speaker's wanting to know whether some relevant property holds of the referent of the description, and 2) the speaker's perhaps wanting that property to hold. Allen and Perrault (1980) suggest that properties needed to "fill in" such fragments come from shared expectations (not just from prior syntactic forms, as is current practice in computational linguistics). The property in question in our domain is IDENTIFIED-REFERENT(HEARER, NP), which is (somehow) derived from the nature of the task as one of manual assembly. Thus, expectations have suggested a starting point for an inference chain — it is shared knowledge that the speaker wants to know whether IDENTIFIED-REFERENT(HEARER, NP). In the same way that questioning the completion of an action can convey a request for action, questioning IDENTIFIED-REFERENT conveys a request for IDENTIFY-REFERENT (see Case 3, Group A, above). Thus, by our positing an IDENTIFY-REFERENT act, and by assuming such an act is expected of the user, the inferential machinery can derive the appropriate intention behind the use of a noun phrase fragment.

The theory should account for 48% of the

identification requests in our corpus, and should be extended to account for an additional 6%. The next group of utterances cannot now, and perhaps should not, be handled by a theory of communication based on reasoning about action.

#### C. Indirect Requests for Confirmation

Group C utterances (as well as Group A, cases 1, 2, and 4) can be interpreted as requests for identification by a rule stipulated by Labov and Fanshel (1977) — if a speaker ostensibly informs a hearer about a state-of-affairs for which it is shared knowledge that the hearer has better evidence, then the speaker is actually requesting confirmation of that state-of-affairs. In Telephone (and Teletype) modality, it is shared knowledge that the hearer has the best evidence for what she "has", how the pieces are arranged, etc. When the apprentice receives a Group C utterance, she confirms its truth perceptually (rather than by proving a theorem), and thereby identifies the referents of the NP's in the utterance.

The indirect request for confirmation rule accounts for 66% of the identification request utterances (overlapping with Group A for 35%). This important rule cannot be explained in the theory. It seems to derive more from properties of evidence for belief than it does from a theory of action. As such, it can only be stipulated to a rule-based inference mechanism (Cohen, 1979), rather than be derived from more basic principles.

#### D. Nearly Direct Requests

Group D utterance forms are the closest forms to direct requests for identification that appeared, though strictly speaking, they are not direct requests. Case 1 mentions "Look on", but does not indicate a search explicitly. The interpretation of this utterance in Perrault and Allen's scheme would require an additional "body-action" inference to yield a request for identification. Case 2 is literally an informative utterance, though a request could be derived in one step. Importantly, the frequency of these "nearest neighbors" is minimal (3%).

#### E. Summary

The act of requesting referent identification is nearly always performed indirectly in Telephone mode. This being the case, inferential mechanisms are needed for uncovering the speaker's intentions from the variety of forms with which this act is performed. A plan-based theory of communication augmented with a rule for identifying indirect requests for confirmation would account for 79% of the identification requests in our corpus. A hierarchy of communicative acts (including their propositional content) can be used to organize derived rules for interpreting speaker intent based on utterance form, shared knowledge and shared expectations (Cohen, 1979). Such a rule-based system could form the basis of a future pragmatics/discourse component for a speech understanding system.

#### IV. RELATIONSHIP TO OTHER STUDIES

These results are similar in some ways to observations by Ochs and colleagues (Ochs, 1979; Ochs, Schieffelin, and Pratt, 1979). They note that parent-child and child-child discourse is often comprised of "sequential" constructions — with separate utterances for securing reference and for predication. They suggest that language development should be regarded as an overlaying of newly-acquired linguistic strategies onto previous ones. Adults will often revert to developmentally early linguistic strategies when they cannot devote the appropriate time/resources to planning their utterances. Thus, Ochs et al. suggest, when competent speakers are communicating while concentrating on a task, one would expect to see separate utterances for reference and predication. This suggestion is certainly backed by our corpus, and is important for computational linguistics since, to be sure, our systems are intended to be used in some task.

It is also suggested that the presence of sequential constructions is tied to the possibilities for preplanning an utterance, and hence oral and written discourse would differ in this way. Our study upholds this claim for Telephone vs. Teletype, but does not do so for our Written condition in which many requests for identification occur as separate steps. Furthermore, Ochs et al.'s claim does not account for the use of identification requests in Teletype modality after prior referential miscommunication (Cohen, 1981). Thus, it would seem that sequential constructions can result from (what they term) planned as well as unplanned discourse.

It is difficult to compare our results with those of other studies. Chapanis et al.'s observation that voice modes are faster and wordier than teletype modes certainly holds here. However, their transcripts cannot easily be used to verify our findings since, for the equipment assembly problem, their subjects were given a set of instructions that could be, and often were, read to the listener. Thus, utterance function would often be predetermined. Our subjects had to remember the task and compose the instructions afresh.

Grosz' (1977) study also cannot be directly compared for the phenomena of interest here since the core dialogues that were analyzed in depth employed a "mixed" communication modality in which the expert communicated with a third party by teletype. The third party, located in the same room as the apprentice, vocally transmitted the expert's communication to the apprentice, and typed the apprentice's vocal response to the expert. The findings of finer-grained and indirect vocal requests would not appear under these conditions.

Thompson's (1980) extensive tabulation of utterance forms in a multiple modality comparison overlaps our analysis at the level of syntax. Both Thompson's and the present study are primarily concerned with extending the

habitability of current systems by identifying phenomena that people use but which would be problematic for machines. However, our two studies proceeded along different lines. Thompson's was more concerned with utterance forms and less with pragmatic function, whereas for this study, the concerns are reversed in priority. Our priority stems from the observation that differences in utterance function will influence the processing of the same utterance form. However, the present findings cannot be said to contradict Thompson's (nor vice-versa). Each corpus could perhaps be used to verify the findings in the other.

## V. CONCLUSIONS

Spoken and teletype discourse, even used for the same ends, differ in structure and in form. Telephone conversation about object assembly is dominated by explicit requests to find objects satisfying descriptions. However, these requests are never performed directly. Techniques for interpreting "indirect speech acts" thus may become crucial for speech understanding systems.

These findings must be interpreted with two cautionary notes. First, the request-for-identification category is specific to discourse situations in which the topics of conversation include objects physically present to the hearer. Though the same surface forms might be used, if the conversation is not about manipulating concrete objects, different pragmatic inferences could be made.

Secondly, the indirection results may occur only in conversations between humans. It is possible that people do not wish to verbally instruct others with fine-grained imperatives for fear of sounding condescending. Print may remove such inhibitions, as may talking to a machine. This is a question that cannot be settled until good speech understanding systems have been developed. We conjecture that the better the system, the more likely it will be to receive fine-grained indirect requests. It appears to us preferable to err on the side of accepting people's natural forms of speech than to force the user to think about the phrasing of utterances, at the expense of concentrating on the problem.

## ACKNOWLEDGEMENTS

We would like to thank Zoltan Ueheli for conducting the videotaping, and Debbie Winograd, Rob Tierney, Larry Shirey, Julie Burke, Joan Hirschhorn, Cindy Hunt, Norma Peterson, and Mike Nivens for helping to organize the experiment and transcript preparation. Thanks also go to Sharon Oviatt, Marilyn Adams, Chip Bruce, Andee Rubin, Rav Perrault, Candy Sidner, and Ed Smith for valuable discussions.

## VI. REFERENCES

- Allen, J. F., A plan-based approach to speech act recognition, Tech. Report 131, Department of Computer Science, University of Toronto, January, 1979.
- Allen, J. F., and Perrault, C. R., "Analyzing intention in utterances", Artificial Intelligence, vol. 15, 143-178, 1980.
- Chapanis, A., Parrish, R., Ochsman, R. B., and Weeks, G. D., "Studies in interactive communication: II. The effects of four communication modes on the linguistic performance of teams during cooperative problem solving", Human Factors, vol. 19, No. 2, April, 1977.
- Chapanis, A., Parrish, R. N., Ochsman, R. B., and Weeks, G. D., "Studies in interactive communication: I. The effects of four communication modes on the behavior of teams during cooperative problem-solving", Human Factors, vol. 14, 487-509, 1972.
- Cohen, P. R., "The Pragmatic/Discourse Component", in Brachman, R., Bobrow, R., Cohen, P., Klovstad, J., Webber, B. L., and Woods, W. A., "Research in Knowledge Representation for Natural Language Understanding", Technical Report 4274, Bolt, Beranek, and Newman, Inc., August, 1979.
- Cohen, P. R., "The need for referent identification as a planned action", Proceedings of the Seventh International Joint Conference on Artificial Intelligence, Vancouver, B. C., 31-36, 1981.
- Cohen, P. R., and Perrault, C. R., "Elements of a plan-based theory of speech acts", Cognitive Science 3, 1979, 177-212.
- Dore, J., Newman, D., and Gearhart, M., "The structure of nursery school conversation", Children's Language, Vol. 1, Nelson, Keith (ed.), Gardner Press, New York, 1978.
- Grosz, B. J., "The representation and use of focus in dialogue understanding", Tech. Report 151, Artificial Intelligence Center, SRI International, July, 1977.
- Labov, W., and Fanshel, D., Therapeutic Discourse, Academic Press, New York, 1977.
- Mann, W. C., Moore, J. A., Levin, J. A., and Carlisle, J. H., "Observation methods for human dialogue", Tech. Report 151/RR-75-33, Information Sciences Institute, Marina del Rey, Calif., June, 1975.
- Ochs, E., "Planned and Unplanned Discourse", Syntax and Semantics, Volume 12: Discourse and Syntax, Givón, T., (ed.), Academic Press, New York, 51-80, 1979.

Ochs, E., Schieffelin, B. B., and Pratt, M. L., "Propositions across utterances and speakers", in Developmental Pragmatics, Ochs, E., and Schieffelin, B. B., (eds.), Academic Press, New York, 251-268, 1979.

Perrault, C. R., and Allen, J. F., "A plan-based analysis of indirect speech acts", American Journal of Computational Linguistics, vol. 6, no. 3, 167-182, 1980.

Robinson, A. E., Appelt, D. E., Grosz, B. J., Hendrix, G. G., and Robinson, J., "Interpreting natural-language utterances in dialogs about tasks", Technical Note 210, Artificial Intelligence Center, SRI International, March, 1980.

Rubin, A. D., "A theoretical taxonomy of the differences between oral and written language", Theoretical Issues in Reading Comprehension, Spiro, R. J., Bruce, B. C., and Brewer, W. F., (eds.), Lawrence Erlbaum Press, Hillsdale, N. J., 1980.

Sacerdoti, E., "Reasoning about Assembly/Disassembly Actions", in Nilsson, N. J., (ed.), Artificial Intelligence — Research and Applications, Progress Report, Artificial Intelligence Center, SRI International, Menlo Park, Calif., May, 1975.

Sidner, C. L., Bates, M., Bobrow, R. J., Brachman, R. J., Cohen, P. R., Israel, D. J., Schmolze, J., Webber, B. L., and Woods, W. A., "Research in Knowledge Representation for Natural Language Understanding", BBN Report 4785, Bolt, Beranek, and Newman, Inc., Nov., 1981

Sinclair, J. M., and Coulthard, R. M., Towards an Analysis of Discourse: The English Used by Teachers and Pupils, Oxford University Press, 1975.

Thompson, B. H., "Linguistic analysis of natural language communication with computers", Proceedings of COLING-80, Tokyo, 190-201, 1980.

Winograd, T., Understanding Natural Language, Academic Press, New York, 1972.