THE NEED FOR REFERENT IDENTIFICATION AS A PLANNED ACTION*

Philip . Cohen

Bolt Beranel and Newman

ABSTRACT

The paper presents evidence that speakers often attempt to get hearers to identify referents as a separate step in the speaker's plan. Many of the communicative acts performed in service of such referent identification steps can be analyzed by extending a plan-based theory of communication for task-oriented dialogues to include an action representing a hearer's identifying the referent of a description — an action that is reasoned about in speakers' and hearers' plans. The phenomenon of addressing referent identification as a separate goal is shown to distinguish telephone from teletype task-oriented dialogues and thus has implications for the design of speech-understanding systems.

I. INTRODUCTION

A. Overview

This paper is the beginning of an empirical and computational exploration into the pragmatics of reference [8, 9, 15], specifically, into speakers' intentions in uttering descriptions. We shall examine only the most clear cut cases — those in which the speaker intends the hearer to pick out a referent in the world.

The paper presents evidence that speakers often attempt to get hearers to identify referents as a separate step in the speaker's plan. Many of the communicative acts performed in service of such referent identification steps can be analyzed by extending a plan-based theory of communication for task-oriented dialogues [4, 5, 12, 20, 22] to include an action representing a hearer's identifying the referent of a description — an action that is reasoned about in speakers' and hearers' plans. The phenomenon of addressing referent identification as a separate goal is shown to distinguish telephone from teletype task-oriented dialogues and thus has implications for the design of speech-understanding systems.

•This research was supported primarily by the National Institute of Education under contract US-NIE-C-400T76-0116, and also in part by the Defense Advanced Research Projects Agency under contract N000U-77-C-0378, monitored by the Offioe of Naval Research.

B. identification

Searle [23] has claimed that the communicative act of (singular definite) referenoe involves uttering an expression D with the intention that the hearer pick out, or identify the referent of D. From the perspective of a plan-based theory, identifying a referent should be represented as an action in the speaker's plan. As such, It can be reasoned about just as any other act. For example, just as speakers can intend the means by which a hearer performs an action, they may expect and intend for hearers to Identify a referent by performing some combination of sensory and cognitive "actions". A speaker uttering the description "the two salty red ones" may be expecting (and perhaps intending) for the hearer to count, look at, and taste various objects. Speakers design their expression D so that hearers can use D as a "blueprint" for the actions needed to identify the referent (in the world).

Many occurrences of descriptions in task-oriented conversations are not intended to be used in this way. For example, in dialogues with an information booth clerk in a train station [1, 13], patrons uttering "the 3:15 to Montreal?" are not intending the clerk to find the train. Instead, as part of their plan for boarding a train, patrons are intending the clerk to given them a co-referring noun phrase that will allow them to identify the train. The attributive use of definite descriptions [8] is another case in which the speaker has no intention that the hearer identify a referent.

Identification is essentially a search process, the act of searching for something that satisfies the description. We can approximate its definition with the following:

IDENTIFY(agt, D) (where D is a description)

precondition: There is an object x
perceptually accessible to ACT
such that x is the referent of D.
effect: (IDENTIFIED-REF agt D)
means: Some function mapping the description D to some procedure that when executed yields a (perhaps) "direct representation" of the referent of D. That procedure may well incorporate the results of perceiving the world.

What la not clear, of course, la Juat what AGT has to know about D to say he haa identified it.* To give a name to the state of knowledge one la in after having identified the referent of D, we will use (IDENTIFIED-REF agt D). We shall only consider the "basic" case of Identification through perception.

C. The problem areas

There are (at least) six problem areas that need to be addressed.

1. Is there evidence for an IDENTIFY act in speakers' and hearer's plans?
2. If so, why is it planned?
3. What advantages accrue from positing such actions? Do they allow us to develop a uniform analysis for what would otherwise be unrelated phenomena?
4. What signals or communicates a speaker'a intention that a hearer identify the referent of a description?
5. When do speakers explicitly communicate the intention that their hearer identify a referent?
6. How is IDENTIFY defined as an act? In what situations is it appropriate? What changes of "state" does it effect? By what means is it accomplished?

This paper will discuss problem areaa 1, 3, and 5 (and touch on problems 2 and 4) by presenting evidence from task-oriented dialoguea. We hope to show that an IDENTIFY act, coupled with a plan-based theory of speech acts will provide a basis for extending our current ability to model task-oriented dialogue.

II. EVIDENCE FOR AN IDENTIFY ACT

la there evidence for an IDENTIFY act? The answer is a qualified "yes" (naturally). To the extent that philosophers, linguists, and AI researchers map action verba and the objects of imperatives onto underlying "acta", then we need one for IDENTIFY. Utterances such as "Notice the two side outlets on the chamber", or "find the rubber ring shaped like an 0.", which occurred in the transcripts, clearly indicate that the speaker intends the hearer to do something. The term "request" labels any action (linguistic or not) of the speaker that results in the hearer's thinking (for the "right" reasons [6, 10, 20, 23]) the speaker wants and intends him/her to do a particular action. Earlier, we claimed that the speaker's plan contained an IDENTIFY act, to be performed by the hearer, representing a physical search for an object satisfying the description. Any action of the speaker'a that communicates the intention that the hearer identify the referent of a description will be labelled a request to identify. By regarding such utterances as requests, we expect to reap the benefits of theories and process models having to do with the generation and understanding of requests. To show some of these potential beneflta, we shall consider utterances that oocurred in our study.

D. The dialogues

The evidence used will be transcript fragments of task-oriented dialoguea. Following the SRI task-oriented dialogue paradigm [11], 25 adult university students ("experts") were videotaped as they instructed other students ("novices") in assembling a toy water pump. Communication employed any one of five communication modes: face-to-faoe, telephone, teletype, audiotape, and written (of. [19]). Part of the purpose of the study was to investigate the constraints that modality (e.g., teletype) impose on discourse structure. In all dialogues, the novices could manipulate the pieces, while the experts could not. In the face-to-face mode, the expert could see the novloe's aasembly. In all other modea, the expert had his/her own set of pieces but could not assemble them.

Utteranoea in the dialoguea were coded by two assistanta trained by the author. For a discussion of the coding system, reliability statistics, and preliminary results see [7]. The coding allowed for the following speech acts (and many others):

- requesting physical actions
- questioning whether a physical action was completed
- informing that a particular phyalcal action was completed
- requesting that the hearer Identify the referent of a description
- questioning whether identification was completed
- informing that identification was completed

E.THE goAl structure of task-oriented dialoguea

The following goal atructure la appropriate to all modes of our dialogues: For each aasembly goal-state (e.g., (CONNECT MAIN-TUBE AIR-CHAMBER)), the expert haa the following goala:

1. Select an action achieving that goal atate.
2. Get the apprentice to perform that action.
3. Ensure success of the action.

Since we assume the experts are creating and executing plans, goals derived from (2) lead to the expert'a planning and exeouting requests for assembly actions. In face-to-faoe mode, the experts achieve goal (3) through vision. For telephone and teletype dialoguea, apprenttices inform the expert when actions are completed. Occasionally, when the apprentice is taking longer than expected (or in face-to-face dialogues, when the apprentice'a body- is blooking the expert's

•Use of a standard name or rigid designator in the possible worlds framework Implies AGT is so "acquainted" with the referent of D that he could not possibly misidentlfy it. That, however, is too atrong a condition. Ultimately, one would like to relativize identification to a "purpose".

view) experts question whether the requested action has been completed. These speech acts, requesting actions, Informing that actions were completed, and questioning the completion of actions will be seen to pertain as well to the Identification of referents. Before showing that the overall structuring is the same, we will discuss the many ways requests get made and show that those same ways arise for identification.

F. Reouaats to Identify

In this section, we present various types of requests occurring in our dialogues and a set of parallel examples of what we will call requesting identification. Requests for action will be abbreviated by "R(ACT)", and requests for identification will be labelled as "R(ID)". The modality of communication is indicated parenthetically.

Direct Requests

- "Fit the blue cap over the tube end." R(ACT) (Written)
- "Notice the two side outlets on the chamber" R(ID) (Written)

Indirect Requests

- "We want to cover the bottom one with a little blue cover." R(ACT) (Telephone)
- "There's a long cylinder that has a slightly purplish cast to it." R(ID) (Telephone;
- "Now you uh, you can place that uh, the rod in through the, that end of the pump" R(ACT) (face-to-face)
- "- On the table is a small, simple red plug without a hole." R(ID) (Written)
- "2. black flexible ring (skinny donut-like)" R(ID) (Written)
- "do you see small/ three small red pieces?" R(ID) (Telephone)
- "now you have two devices that that are clear plastic" R(ID) (telephone)
- "Now. the smallest of the red pieces?" R(ID) (Telephone)

Many of these examples are discussed in detail below. We have already discussed direct requests to Identify stated as single imperatives. Importantly, these utterances show that the goal of identification can be achieved in a separate step. Because of the goal structure of our task-oriented dialogues, the success of identification requests should be questioned and reported Just like requests for physical acts. We shall return to this point later.

G. Indirect requests to identify

Perhaps the most interesting aspect of requests is that they need not occur as imperatives. If the intentions behind an utterance are to be inferred from its form, then it is problematic that utterances nominally expressing one intention are often used to get the hearer to infer another intention. For example, utterances beginning with "You can <act>" might literally mean "You now have permission to do <act>" — i.e.. are actually informing the hearer of something. However, in our situation, suoh utterances are taken to mean "Do <aot>". Clearly, not all situations in which one person gives another permission to do something allow one to infer the speaker wants the act to be done (now). (Consider "You can go on that trip without me, if you want to.")

For our transcripts, this problem of inferring the right intention occurs even more frequently with identification. Regarding the above examples, "there is . . ." is literally an informing act about the state of the world, as is "On the table is . . .", and "Now you have two . • .". The same utterances could appear in other circumstances and would not lead a listener/reader to infer that an object was intended to be identified. For example, as part of the setting of a mystery story, similar statements, whatever their function, are obviously not intended to get the reader to search his/her perceptual world. Our task, as analysts, is to uncover the conditions under which hearers are to make such inferences. Perrault and Allen [20] show how, by assuming hearers are trying to recognize speakers' plans, a speaker's ostensibly informing the hearer that the precondition to some act is true can be seen as requesting the hearer to perform that act, provided it is shared knowledge that the speaker would want the primary effect of that act.

Assuming that the precondition to IDENTIFY is that there must exist an object perceptually accessible to the hearer satisfying the description, then, for utterances whose logical form is EXISTS (X): D(X), Perrault and Allen's general method applies. The method generates an IDENTIFY act as part of the hearer's model of the speaker's plan, provided it is shared knowledge that the speaker has some reason to want the description to be identified (for example, to enable the hearer to perform a physical action). Thus, the plan-based inference that someone may want the precondition for an act to hold because they want to do that act generalizes to IDENTIFY.

Other evidence to substantiate the generalization from requesting physical acts to requesting IDENTIFY is in the area of questions. Many researchers (e.g., Labov and Fanshel [16]) have noticed that some questions about whether an act (or the final state of an act) has been accomplished convey the speaker's intention that the hearer do that act if she has not already done so. Thus "Is the garbage out?" can be a request to take out the garbage, provided the speaker Is believed to want this state of affairs. Perrault and Allen's model extends easily to this oase.

The same phenomenon appears to occur with "do you see small/three small red pieces?" The hearer may not be looking at the pieces, (but at the camera, the floor, or somewhere else) and could perhaps truthfully answer "no". Of course, the speaker intended the hearer to look for, i.e., IDENTIFY, something described as "the small red piece."

Yet other phenomena that appear susoeptible to the same treatment are certain oases of elliptical questions, such as "Now, the smallest of the red pieces?" The action to be performed is not

explicitly stated here and must be supplied on the basis of shared knowledge about the situation — who can do/see what, what each thinks the other believes, what is expeoted, etc. Such knowledge will be needed to differentiate the intentions behind "the 3:15 to Montreal?", where the description's identification is not Intended, froa those underlying the above sentence, where the hearer is intended to find the pieoe.

According to the theory, the speaker's intentions conveyed by the elliptical question include 1) the speaker's wanting to know whether some property holds of the referent of the description (0), and 2) the speaker's perhaps wanting that property to hold. Allen [1] suggests that properties that need to be inferred to "fill in" fragments come from shared expectations. In our domain, it is shared knowledge that the hearer will pick up pieces that the expert describes (since both know it is a manual assembly task). Given the expectation to pick up, the property in Allen's scheme that must hold is that the hearer KNOWREF D (where KNOWREF is defined as EXIST(X) [X = D] & H BELIEVE [X = D]), which is an exceedingly strong way to say that the hearer has identified the referent. Since this is the first reference to the piece in question, it is shared knowledge that the hearer has not previously identified the referent of D. In the same way that questioning the completion of an action can convey a request for that action, questioning IDENTIFIED-REF can convey a request to IDENTIFY. Notice that the expectation that the description is to be identified is independent of the specific expectation to PICK-UP; it depends only on the expectation that the hearer is to perform some action on a physloally present objeot. For example, while a nuclear reactor-scrubbing robot would expeot commands to perform action(a; on the interior of the reaotor, it should not attribute PICK-UP as a speaker's intention solely on hearing "there's a glowing rod".

H. Evidence *far* IDENTIFY froa Hearer Responses

Another kind of evidence to support our claim that identification is a planned and recognized action is participants' performance of the same speech acts with the same surface forms in response to what we are calling requests for identification as they do in responding to requests for action. For example, in a telephone dialogue we found the following:

a. A: "and attach the pink thing
so that it covers the
hole in the middle."
REQUEST(A,B, MESH(B,VALYE2,TUBE-BASE))

b. B: "Got it
One way valve.
We're all set"
INF0RM(B,A, COMPLETED(B, MESH(...)))

o. A: "Ok, now
there' A, black JteEiAg"
REQUEST(A,B, IDENTIFY(B,"a black o-ring"))

d. B: "Got it"
INF0RM(B,A, COMPLETED(B,IDENTIFY(..)))

Speaker B used "Got it" frequently for signalling that he was finished performing actions. It would therefore be unreasonable to say that "got it" in (c) was suddenly intended literally to convey that B was simply informing that he was holding the piece.

We have shown that, just like requests for action, requests for identification occur as separate steps, come in the same forms, and provoke the same responses froa hearers. We have also suggested how existing computational models might be applied to IDENTIFY and yield appropriate analyses. It should be clear that by positing an IDENTIFY act, coupled with a plan-based account of indirect speech act interpretation, we stand to gain a more unified account of apparently similar phenomena that would otherwise be unrelated.

III. WHEN ARE SEPARATE REQUESTS TO IDENTIFY USED?

Major questions for computational models of language production and comprehension become:

- How and why should a system plan separate requests for identification rather than simply produce imperatives requesting assembly actions?
- 0 Under what conditions will systems be confronted with such utterances?

I. Production

The answer to the first question is suggested by the following fragment of a teletype dialogue:

1. B: "anyway, put the red piece with the strange projections LOOSELY into the bottom hole on the main tube. Ok?"
2. A: "Which hole the bottom one on the side?"
3. B: "right put the 1/4 inch long 'post' Into the loosely fitting hole...
4. N: I don't understand what you mean
5. B: the red piece, with the four tiny projections?
6. N: OK
7. B: Just place it loosely [into the]
8. N: [done][B & N typed simultaneously, causing gibberish to appear on each screen]
9. B: yes?
10. N: yes
11. B: place it loosely into the hole on the side of the large tube...
12. N: done
13. B: very good. See the clear elbow tube?
14. N: Yes
15. B: Place the large end over that same place.
16. N: ready

17* B: take the clear dome and attach it to the end of the elbow Joint...

18. N: using the blue attachment part?

19. B: right, it's already attached, so I didn't mention it Now, put the red nozzle over the hole in the dome."

There are three strategies of instruction here. First, direct requests for assembly actions, in the form of imperatives, as in line (1). Second, there are conjoined direct requests, for picking up followed by an assembly action as in (17). Finally, B performs separate requests to identify, each followed by a request for an assembly action as in (5) - (7) and (13) - (15).

What is important to notice here la that B shifts his strategy (in a fashion that resembles driving a three-speed oar). Prior to this fragment, the conversation had proceeded smoothly, in "high gear", so to speak, with B initially "upahifting" from first a "take and assemble" request to six oonsecutlve assembly requests (one of them indirect), the last of which is utterance 1 of this fragment.

In (2)-(ii), we observe clarification dialogue about a prior noun phrase. Immediately after an apparent breakdown at (4), B "downshifts" to questioning the achievement of his first subgoal, identifying the red piece. Once that is corrected, and a channel contention problem is solved, B stays in "low gear", explicitly ensuring success of his reference (in (13)-(14)) before requesting an assembly action in (15). After that success, he "upshifts" to "second gear" — with requests to pick-up and assemble in (17). After being successful yet again, B "upshifts" in (19) to "high gear", again using a dlreot assembly request, and stays In "high" for the rest of the dialogue (six more requests).

What could explain this conversation pattern? A representation of the plan for assembling shows clearly that to install a piece, one must be holding it; to hold it, one must pick it up; to perform any action on an object, one must have identified that object. By requesting an assembly action ("high gear"), one requires the listener to make the remaining inferences. By requesting the sequence take-and-aasemble ("second gear"), the speaker makes one of the inferences himself, but requires the listener to realize that identification of the speaker's part description is needed. Finally, "low gear" involves the speaker's checking the success of the component subgoals, which involves description identification.

In summary, the strategy shift to "low gear" occurs after a referential miscommunlcatlon because it affords a more precise monitoring of the listener's achievement of goals.* A task-oriented dialogue system should have "dialogue sense" enough to plan utterances tailored to the user's problem.

J. Language Comprehension — Speech vs. Teletype

Our second question is "Under what conditions j will systems be confronted with such utterances and speech acts?" We have observed that REQUEST(IDENTIFY) and QUESTION(IDENTIFIED-REF) occur frequently in telephone communication and quite Infrequently in teletype communication. The magnitude differences in our transcripts are quite striking. For example, the only examples of either speech act occurring in five teletype dialogues are the two above — where the speaker is concentrating on identification because of (we suspect) prior referential misoommunication. By way of contrast, these speeoh acts are prevalent, if not predominant, in all five telephone dialogues. The total number of instances of these speeoh act types ranges from a low of seven to a high of 25 (mean = 15.7) per dialogue.

IV. CONCLUDING REMARKS

We have Justified the need for a planned IDENTIFY action with evidence from human dialogues. By incorporating an IDENTIFY act into a plan-based theory of speech acts, the resulting theory can explain regularities of dialogue for which It was not designed. In particular it accounts for speakers' requesting that hearers identify the referents of descriptions, for their questioning the success of referent identification, and for hearers' reports of that success. We have suggested how computational methods ourrently being used to "reason out" the oonveyed force of a class of indirect requests for action can be applied successfully to a class of (what we will call) indirect requests for identification.

Analysis of teletype and telephone interactions shows a marked difference across modes in the use of explicit requests for identification. Thus, speech understanding systems may have to be prepared for language use that departs quite dramatically from that addressed to teletype-based systems. We speculate that the more competent the speech-understanding system is, the more it will receive such requests. We expect these findings to also have ramifications for the design of task-oriented dialogue systems [21, 3], for distributed artificial intelligence systems [14], and for models of language generation [2, 17].

There are obviously many unsolved problems regarding identification. We do not know how to define such an act, how to construct descriptions to ensure its success nor how it becomes a goal. Nevertheless, in this paper, we hope to have shown the necessity and utility of considering identification in a model of natural language processing for task-oriented dialogue. However our data are to be explained, we expect identification to play part.

V. ACKNOWLEDGEMENTS

Many thanks to Scott Fertig, Kathy Starr, and Larry Shirey for their invaluable discourse analyses, to Debbie Winograd for aid in conducting the experiments, to Rusty Bobrow, Chip Bruce, David Israel, Ray Perrault, and Candy Sidner for Influential discussions and critloal reading, and to Norma Peterson for outstanding assistance in preparing this paper.

•The use of separate utterances for identification has also been observed independently by Ochs, Schieffelln, and Pratt [18] for parent-child discourse.

VI. REFERENCES

1. Allen, J. F. A plan-based approach to speech act recognition. Technical Report 131, Department of Computer Science, University of Toronto, January, 1979.

2. Appelt, D. Problem-solving applied to language generation. Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, 1980.

3. Brachman, R., Bobrow, R., Cohen, P., Klovstad, J., Webber, B. L., & Woods, W. A. Research in natural language understanding. Technical Report 4274, Bolt Beranek and Newman Inc., August, 1979.

4. Bruce, B. C. Belief systems and language understanding. Technical Report 2973, Bolt Beranek and Newman Inc., 1975.

5. Cohen, P. R., & Perrault, C. R. Elements of a plan-based theory of speech acts. Cognitive Science 3, 1979, 177-212.

6. Cohen, P. R. and Levesque, H. L. Speech Acts and the Recognition of Shared Plans. Proc. of the Third Biennial Conference, Canadian Society for Computational Studies of Intelligence, Victoria, B. C., May, 1980, 263-271.

7. Cohen, P. R., Fertig, S., Shirey, L., Starr, K., & Tierney, R. Pragmatics and the modality of communication. In preparation

8. Donnellan, K. Reference and definite description. The Philosophical Review 75, 1960, 281-304. Reprinted in Steinberg & Jacobovits (Eds.), Semantics, Cambridge University Press, 1966.

9. Donnellan, K. S. Speaker Reference, Descriptions and Anaphora. In Syntax and Semantics: Volume 9, Pragmatics, Cole, P., (Ed.), New York, 1978, 47-68.

10. Grice, H. P. Meaning. Philosophical Review 66, 1957, 377-388.

11. Grosz, B. J. The representation and use of focus in dialogue understanding. Technical Report 151, Artificial Intelligence Center, SRI International, July, 1977.

12. Hobbs, J., & Evans, D. Conversation as planned behavior. Cognitive Science 4, 4, 1980, 349-377.

13. Horrigan, M. K. Modelling simple dialogues. Technical Report 108, Department of Computer Science, University of Toronto, 1977.

14. Konolige, K., & Nilsson, N. J. Multiple-agent planning systems. Proceedings of the First Annual National Conference on Artificial Intelligence, August, 1980.

15. Kripke, S. Speaker's reference and semantic reference. In Midwest Studies in Philosophy, French, P. A., Uehling, T. E., Wettstein, H. K., (Eds.), Minneapolis, 1977, 255-276.

16. Labov, W., & Fanshel, D. Therapeutic discourse. Academic Press, New York, 1977.

17. McDonald, D. D. Natural Language Production as a Process of Decision-making Under Constraint. Ph.D. Th., M. I. T., Cambridge, MA, August 1980.

18. Ochs, E., Schieffelin, B. B., & Platt, M. L. Propositions across utterances and speakers. In Developmental Pragmatics, Ochs, E., & Schieffelin, B. B., (Eds.), New York, 1979.

19. Ochsman, R. B, and Chapanis, A. The effects of 10 communication modes on the behaviour of teams during co-operative problem-solving. International Journal of Man-Machine Studies 6, 5, 1974, 579-620.

20. Perrault, C. R., & Allen, J. F. A plan-based analysis of indirect speech acts. American Journal of Computational Linguistics 6, 3, 1980, 167-182.

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

22. Schmidt, C. F. Understanding human action. Proceedings of Conference on Theoretical Issues in Natural Language Processing, Cambridge, Massachusetts, 1975.

23. Searle, J. R. Speech acts: An essay in the philosophy of language. Cambridge University Press, Cambridge, 1969.