The Pragmatics of Referring and the Modality of Communication[1]

Philip R. Cohen

Laboratory for Artificial Intelligence Research
Fairchild Camera and Instrument Corporation
Palo Alto, CA

This paper presents empirical results comparing spoken and keyboard communication. It is shown that speakers attempt to achieve more detailed goals in giving instructions than do users of keyboards. One specific kind of fine-grained communicative act, a request that the hearer identify the referent of a noun phrase, is shown to dominate spoken instruction-giving discourse, but is nearly absent from keyboard discourse. Most important, these requests are only achieved "indirectly". – through utterances whose surface forms do not explicitly convey the speakers' intent. A plan-based theory of communication is shown to uncover the speakers' intentions underlying many cases of indirect identification requests found in the corpus, once an action for referent identification has been posited. In so doing, the theory demonstrates how intent (or plan) recognition can be applied in reasoning about the use of a description. As a consequence of this approach, it is shown that the conditions on the planning of successful identification requests account for Searle's conditions on the act of referring. It is concluded that intent recognition will need to be a central focus for pragmatics/discourse components of future speech understanding systems, and that computational linguistics needs to develop formalisms for reasoning about speakers' use of descriptions.

1 Introduction

As natural language interaction with computers becomes more widespread, systems' abilities to engage users in discourse will become increasingly important. These capabilities will be especially in demand when users can speak naturally to their machines. Although it is widely suspected that spoken language is different from written language, the question of precisely what the differences are has only recently become a topic of computational linguistic research. Previous investigations have concentrated on syntactic differences between spoken and written language (Hindle 1983, Kroch and Hindle 1982, Thompson 1980), with the goal of adapting parsing techniques to handle the syntax of spoken language. However, even if this goal were achieved, a system needs to be prepared to handle any unique properties of the *discourse structure* of spoken interaction if it is to be successful in conducting a dialogue.

Of course, there has been much work on discourse processing within computational linguistics (e.g., Grosz 1977, Sidner 1979, Webber 1978), and any future systems will undoubtedly incorporate previously successful techniques. However, one suspects that the coverage

[1] This research was supported primarily 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., and in part by the Oregon State University. The paper was written at the Fairchild Camera and Instrument Corporation, and revised at SRI International. The revisions of this paper have been made possible by a gift from the System Development Foundation to SRI as part of a coordinated research effort with the Center for the Study of Language and Information, Stanford University. Author's current address: Artificial Intelligence Center, SRI International, Menlo Park, CA 94035.

0362-613X/84/020097-50$03.00

of discourse processing algorithms may depend on the corpora from which they were developed, and many of those reflect keyboard-mediated dialogue. Thus, to determine whether and how current techniques need to be adapted to the way people speak, research is needed to compare the discourse structure of spoken and keyboard interaction.

This paper presents an empirical study and theoretical analysis of utterance form and function as determined by the communication modality. The two initial objectives are:

1. To develop an empirical methodology for analyzing discourse pragmatics.
2. To use that methodology to identify both the goals that speakers attempt to achieve in spoken and keyboard modalities, and the discourse and sentence structures they use in achieving those goals.

These objectives are investigated in a study of instruction-giving discourse, a communication task that intimately ties utterance function to a nonlinguistic task being accomplished by the conversants. In addition to depending on the communication task, the goals people achieve with language are a function, in part, of the communication situation – i.e., of how the speaker(s), hearer(s), object(s) under discussion, and discourse itself are situated in the world. For example, the utterance "On the table is a little yellow piece of rubber," would be interpreted quite differently in a narrative than in a set of instructions for assembling an object. The communication situation helps to determine the *pragmatics of reference* – what speakers intend hearers to do with referring expressions. Thus, a third goal of this paper, a subsidiary to objective 2 is to consider

3. How the speakers' goals for the interpretation of referring expressions are expressed and achieved in different modalities.

Results indicate that speakers attempt to achieve more detailed referential goals in giving instructions than do users of keyboards. That is, speakers explicitly request hearers to identify the referents of noun phrases (NPs), but users of keyboards do not. Instead, the referential goals achieved by these requests are subsumed by other requested actions. Most importantly, these identification requests are only achieved "indirectly" – through utterances whose surface forms do not explicitly convey the speakers' intent.

Current theories propose that the speaker's intentions underlying the use of indirect speech acts can be recognized as a by-product of a more general, independently motivated process of inferring a speaker's *plans* (Bruce 1983, Cohen and Perrault 1979, Cohen and Levesque 1980, Perrault and Allen 1980, Schmidt 1975, Sidner and Israel 1981). Essentially, illocutionary acts, which communicate the speaker's intentions, are regarded as steps in a speaker's plan, just as physical acts are. Furthermore, just as observing an agent's behavior may lead one to infer what the agent is trying to do, so too can the observation/understanding of a speaker's utterance lead an observer to infer the speaker's intentions. This approach has led to formal and computational models of discourse processing (Allen 1979; Allen and Perrault 1980; Brachman et al. 1979; Cohen and Levesque, in preparation; Sidner et al. 1981). Although these provide a more comprehensive account of indirect speech act interpretation than previous linguistic or philosophical approaches, they have not been tested against a corpus other than the ones that supported their creation (e.g., Horrigan 1977). Therefore, as an adequacy test, the fourth objective for this paper is:

4. To evaluate how well a plan-based theory of communication can uncover the intentions underlying the use of many surface forms in the transcripts.

The theory is shown to account for approximately 70% of the indirect requests for referent identification found in the transcripts, once an action for referent identification has been posited. An important aspect of the account is the demonstration that speakers and hearers can reason about referent identification much as they reason about other actions and plans. Hence, the last goal for this paper is to

5. Contrast the plan-based analysis of referring, and the flexibility it allows, with Searle's account of reference as a speech act.

I show that Searle's analysis cannot account for many of the examples treated here, and that those examples it does cover can also be handled by the present analysis.

The conclusions I draw are specific to the conversational task of giving instructions about objects physically present to the hearer. This task was chosen for four reasons:

- First, it was expected that speakers would frequently issue requests. Because requests dominate interactions with many question-answering systems, and with most conceivable interactive applications of natural language processing, they have been extensively studied in computational linguistics.
- Second, because the task is simple and constrained, it provides an excellent adequacy test for proposed theories and computational techniques; any theory of communication that cannot handle the phenomena of this study can hardly be called general. However, since the domain is functionally similar to those of various keyboard-based systems (Brachman et al. 1979, Robinson et al. 1980, Winograd 1972), the data and results of the study may suggest directions for extending those systems.
- Third, the domain is similar to those analyzed by other researchers (Chapanis et al. 1972, Chapanis et al. 1977, Grosz 1977), and thus the dialogues could serve to confirm or refute their results.

- Finally, instructions play a crucially important role in people's everyday lives – success in industrial, academic, and bureaucratic tasks, for example, requires the following of instructions. Children are initially instructed face-to-face, but they eventually learn to follow written instructions. The present study, though not this paper, should ultimately provide a window on how the language of written instructions differs from that of instructional dialogue.

In summary, this paper applies an empirical methodology for analyzing discourse pragmatics to compare spoken and keyboard language for an instruction-giving task. The primary differentiating phenomenon, the use of explicit identification requests, is used as an adequacy test for a plan-based theory of communication. Conversely, the theory is used to explicate how such communicative actions might be analyzed. Finally, the theory gives rise to a pragmatic analysis of referring that subsumes Searle's (1969).

The remainder of this paper is organized as follows: Section 2 discusses previous research in a number of related areas. Section 3 isolates the phenomena of interest – referent identification. Section 4 describes the study and the method of discourse analysis, and Section 5 presents and discusses the empirical results. Section 6 sketches the plan-based theory of communication and applies the theory to key examples. Section 7 counters possible alternative explanations that would purport to explain the data without recourse to an analysis of speaker intent. Section 8 compares the analysis with Searle's, and suggests generalizations based on independently motivated principles. Finally, the appendices contain the materials, transcripts, and codings on which these analyses are based. A full set of coded transcripts may be obtained on request.

2 Previous Research

Four traditions of research bear on the problems at hand: empirical work on differences between spoken and written language, discourse analysis, psychological studies of referential communication, and computational linguistics studies of discourse that have been based on observation of actual communicative interaction.

The subject of oral/written language comparisons has received much attention from researchers: Anthropologists have traditionally studied characteristics of "oral" and "literate" cultures; human-factors researchers have investigated the opportunities that particular modalities afford for effective communication; and educational psychologists, using empirical and anthropological methods, have sought answers to children's reading and writing problems in the study of oral/written language differences.

Rubin (1980) discusses methodological weaknesses in many oral/written studies – weaknesses that stem from a simplistic division of language experiences into "oral" and "written". Instead, she classifies language experiences in terms of their characteristic values on several dimensions such as: the use of voice or print, the ability of "speaker" and "hearer" to interact, their spatial and/or temporal commonality, their mutual involvement in the discourse, and the concreteness of the referents. Face-to-face conversations about physically present objects are seen to lie at one extreme within this "communication space" (with "positive" values on the above dimensions), whereas written text is at the opposite. Other language experiences that differ along these dimensions include communication by telephone, keyboard, audiotape, picturephone, writing, etc.[2] Rubin reports that many studies comparing language experiences present conclusions about oral/written language differences even though the language experiences differ from one another along multiple dimensions. In such cases it is not clear if the observed differences result, for example, from the presence of voice, the ability to interact, or both.

There is evidence that at least some *quantitative* linguistic and efficacy results are primarily determined by the presence of voice in the communication modality. A series of studies by Chapanis and colleagues (Chapanis et al. 1972, Chapanis et al. 1977) compared problem-solving effectiveness among teams communicating in face-to-face, voice only, written, keyboard, and other communication modalities. Dependent measures included problem solution time, number of words, sentences, utterances, etc. Results indicate that problems are solved twice as fast in vocal modalities as they are in written ones, even though communicators use twice as many words when speaking.[3] Although motivating the development of speech-understanding systems, these results unfortunately tell us little about how the processing of spoken utterances differs from the processing of written ones.

Other research has compared the syntax of spoken and written discourse. The primary findings are: Written language is syntactically more integrated than spoken, employing nominalizations, participles, complements, relative clauses, etc. (Chafe 1982); and spoken language exhibits regular patterns of false starts and hesitations (Hindle 1983, Kroch and Hindle 1982). The former results can help a system designer to determine which syntactic constructs to emphasize in a grammar for parsing. The latter results are more useful to computational

[2] This approach essentially characterizes language situations as multidimensional vectors whose components, describing the above dimensions, are binary values (e.g., +/- voice). Thus, it is assumed that neighboring modalities afford equal communicative possibilities in all dimensions in which they are the same. This is obviously untrue for the dimension of interaction.

[3] Thompson (1980) has confirmed Chapanis et al.'s results for face-to-face and keyboard modalities.

linguistics. Not only are regularities in nongrammatical speech identified, but a class of "editing" rules is provided that can make such utterances parsable. Current work on relaxing grammar rules and on parsing ill-formed input (Hayes and Mouradian 1981, Kwasny and Sondheimer 1981, Weischedel and Black 1980) is in much the same spirit.

The purposes at hand require analyses of the pragmatic and discourse structure of actual dialogues. Grosz (1977) and Bruce (1981) (among others) have shown how such discourse analyses can have direct implications for algorithm design. In their work, transcripts of dialogues were collected and analyzed, leading to the development of algorithms for speech-understanding systems (Walker 1978, Woods et al. 1976).[4] Grosz' analyses indicate that anaphoric reference in task-oriented dialogues is constrained by the hierarchical structure of the physical task. A parallel structuring of "focus spaces" was proposed as a mechanism to constrain the search for co-referents, and became the mainstay of the discourse component of two systems (Robinson et al. 1980, Walker 1978). Although this research did not directly address the problem of discovering cross-modal similarities and differences, the major finding of explicitly "stacked" topics serving to constrain co-reference was validated independently in a domain of casual, face-to-face conversation (Reichman 1981).

Bruce's pragmatics component for the HWIM system was based on transcripts of human keyboard-mediated dialogues simulating interactions with a travel budget manager. Users were seen to be interacting in various "modes" (e.g., editing-a-trip mode, creating-a-trip mode, etc.). The system attempted to track the user's progress through these modes, using an ATN-based representation, and thereby to create expectations of his/her future utterances. Discourse analysis revealed that users did not follow the strict embedding of subdialogues required by the ATN model. Consequently, the pragmatics component was reorganized as a "demand" model in which the system was seen as responding to one of a set of pending goals. Although this research did not directly address issues of cross-modal similarities and differences, it did point out the promise of a goal-oriented view of language processing.

Many researchers in the field of discourse analysis have tried to identify goals or intentions in dialogue. For example, Labov and Fanshel (1977) analyzed transcripts of therapy sessions by employing the vocabulary of linguistics and speech act theory. Their analyses presented rules for interpreting the intentions behind utterances of various syntactic forms – e.g., rules for when a hearer will interpret utterances as indirect requests for physical action or verbal confirmation. However, these rules were stipulated as regularities of discourse rather than as derived from underlying processes. Their findings should serve as data to be explained, rather than as a satisfying account of discourse.

In research more relevant to computational linguistics, Mann et al. (Mann, Moore, and Levin 1977; Mann, Carlisle, Moore, and Levin 1977) applied traditional empirical methods to the identification of speaker intention and utterance function in dialogue.[5] Their goal was to build systems to replicate observers' scorings of transcripts. The observers, and ultimately the systems, were to identify repeated reference, requests, expressions of comprehension, topic structure, etc., in keyboard dialogues between a user and a computer operator, and in radio dialogues between Apollo astronauts and ground control. Much care was taken to develop a scoring scheme, train dialogue observers, and attain reliability Mann, Carlisle, Moore, and Levin 1977). A separate computer program was to have been built for processing each transcript. By merging the common features of these systems, an empirically-based theory and computational model were to have been developed. This work resulted in a goal-directed, "dialogue games" model of conversational interaction (Levin and Moore 1977), though it is not clear whether the model's formulation resulted from the merging of implementations.

Finally, there is a huge literature of psychological studies of referential communication. I will not survey it here (but see Dickson 1981 for recent papers and Asher 1979 for an extensive review), but mention only two themes of relevance to this study. First, such work has shown that, in spoken interaction, noun phrase length tends to decrease as subsequent references to an object are made. However, in non-interactive spoken modalities (Krauss & Weinheimer 1966), the decrease for subsequent references is lessened. These results indicate that efficiency in referential communication is a function of user feedback.

The development of the component skills involved in referring is a second theme in this literature. In order to test Piaget's "egocentrism" hypotheses, a typical question asked is whether children take their listener's "perspective" into account when planning their referring expressions.[6] Another question raised is whether children of certain ages can adequately make comparisons of the properties of referents and non-referents in order to formulate an adequate referring expression. This line of

[4] However, neither corpus incorporated true spoken interaction. The SRI dialogues that were analyzed in depth were taken from a mixed communication mode in which one, an "expert", typed instructions to a third party, who spoke them to an "apprentice", and typed the apprentice's spoken replies to the expert. The BBN "incremental simulation" dialogues involved only keyboard communication.

[5] Similar approaches include those of Dore et al. (1978), and Sinclair and Coulthard (1975).

[6] Shatz and Gelman (1973) showed they can do so (though not necessarily accurately (Asher 1979)) at a much earlier age than had been supposed.

work is more relevant to the present concerns of characterizing the act of identification, but the subskills examined are still too coarse for our needs.

Previous research has thus provided many lessons, among them:

- the need to compare (at least initially) modalities that are minimally different,
- the need for repeatable methods for characterizing linguistic behavior at the pragmatics and discourse structure level,
- the need to assess the adequacy of our theories, and
- the need for couching explanations in computational terms.

The present study addresses each of these needs.

3 The Phenomena of Interest: Referent Identification

One referential goal that is essential to the present communication task is to get the hearer to identify the object the speaker has in mind. I shall be using the term "identify" in a very narrow, though important and basic, sense – one that intimately involves perception. Thus, the analysis is not intended to be general; it applies only when the referents are perceptually accessible to the hearer, and when the hearer is intended to use perceptual means to pick them out. For the time being, I shall explicitly not be concerned with a hearer's mentally "identifying" some entity satisfying a description, or discovering a co-referring description, although these operations are certainly important aspects of processing many referring expressions. In the remainder of this section, properties of the referent identification act are examined, in part by contrasting it with other concepts that have previously entered into computational linguistic analyses of reference.

Referent identification requires an agent and a description. The essence of the act is that the agent pick out the thing or things satisfying the description. The agent need not be the speaker of the description, and indeed, the description need not be communicated linguistically, or even communicated at all. A crucial component of referent identification is the act of perceptually searching for something that satisfies the description. To determine which method(s) should be used in identifying the referent, the agent first requires some representation of the description *per se.* The description is decomposed by the hearer into a *plan* of action for identifying the referent. The intended and expected physical, sensory, and cognitive actions to be included in that plan may be signalled by the speaker's choice of predicates. For example, a speaker who utters, "the magnetic screwdriver, please", may expect and intend for the hearer to place various screwdrivers against some piece of iron to determine which is magnetic. Similarly, a speaker uttering the description "the three two-inch long salted green noodles" may expect and intend the hearer to count, look at, measure, and perhaps taste various objects. For their part, hearers decompose the noun phrase/description to discover that "green" is determinable by vision, "inch" by measuring, "salted" primarily by taste, "noodle" primarily by vision, and "three" by counting. Speakers know this is what hearers can do, and thus, using a model of the hearer's capabilities and the causal connections among people, their senses, and physical objects, design the referring expression D to suggest the actions needed to identify the referent.

Speakers often not only plan for hearers to identify the referents of descriptions, but also communicate, in the Gricean way (1957), their intention that the hearers do so. This intention may not be explicitly signalled in the utterance, but rather have to be recognized by the hearer. To respond appropriately, a hearer decides when identification is the intended act to perform in response to a description, what part this act will play in the speaker's and hearer's plans, and when to perform the act. If perceptually identifying a referent is represented as an action in the speaker's plan, hearers could reason about it just as they do about any other act, thereby becoming able to infer the speaker's intentions behind, for example, indirect identification requests.

3.1 A sketch of a definition of perceptual referent identification

Figure 1 presents a sketchy definition of the referent identification action, in which the description is formed from "a/the y such that D(y)".[7]

$$
\begin{array}{l}
\forall\ \mathrm{D\ Agt} \\
\quad \exists\ \mathrm{X\ [PERCEPTUALLY\text{-}} \\
\qquad \mathrm{ACCESSIBLE(X,\ Agt)}\ \& \\
\qquad\qquad \mathrm{D(X)}\ \& \\
\qquad \mathrm{IDENTIFIABLE(Agt,D)]} \\
\qquad\qquad \supset \\
\quad \exists\ \mathrm{X\ [RESULT(Agt,} \\
\qquad \mathrm{IDENTIFY\text{-}REFERENT(D),} \\
\qquad\quad \mathrm{IDENTIFIED\text{-}REFERENT} \\
\qquad\qquad \mathrm{(Agt,\ D,\ X)]}
\end{array}
$$

Figure 1. *The act of referent identification.*

[7] This definition is not particularly illuminating, but it is not any vaguer than others in the literature, including Searle's (1969). The point of giving it is that if a definition can be given in this form (i.e., as an action characterizable in a dynamic logic), a plan-based analysis (see section 7.5) applies.

The formula follows the usual axiomatization of actions in a dynamic logic: **P** ⊃ [Act]Q; that is, if P is true, after doing Act, Q holds. Following Moore's (1980) possible worlds semantics for action, the modal operator RESULT is taken to be true of an agent, an action, and a formula, iff in all world states resulting from the agent's performing that action, the formula is true.[8]

The antecedent says there exists some (perhaps more than one) object satisfying three conditions. The first is a "perceptual accessibility" condition to guarantee that the **IDENTITY-REFERENT** action is applicable. This should guarantee that, for example, a speaker does not intend someone to pick out the referent of "3", "democracy", or "the first man to land on Mars". The condition is satisfied in the experimental task because it rapidly becomes mutual knowledge that the task requires communication about the objects in front of the hearer.

The second condition states that **X** fulfills the description **D**. Here, I am ignoring cases in which the description is not literally true of the intended referent, including metonymy, irony, and the like (but see Perrault and Cohen 1981). Finally, **D** should be a description that is identifiable to this particular **Agt**. It should use descriptors whose extension the agent already knows or can discover by action. I am assuming that we can know that a combination of descriptors is identifiable without having formed a plan for identifying the referent.

If the antecedent is true, then the agent picks out something (not necessarily the object satisfying the antecedent) as the referent of **D**. His picking out the "right" (i.e., the intended) object is handled by a separate characterization of the speaker's intention with respect to this action (see section 7.5). Here, I will merely give a name to the state of knowledge the agent is in after having identified the referent of **D** – **(IDENTIFIED-REFERENT Agt D X)**. That is, **Agt** has identified the referent of **D** to be **X**. Of course, what has been notoriously difficult to specify is just what **Agt** has to know about **X** to say he has identified it as the referent of **D**. Clearly, "knowing who the **D** is" (Hintikka 1969, Moore 1980) is no substitute for having identified a referent. After having picked out the referent of a description, we may still not not know who the **D** is. On the other hand, we may know who or what the description denotes, for example, by knowing some "standard name" for it, and yet be unable to use that knowledge to pick out the object. For example, if we ask "Which is the Seattle train?" and receive the reply "It's train number 11689", we may still not be able to pick out and board the train if its serial number is not plainly in view. Clearly, the notion of identification needs to be made relative to a purpose, which perhaps could be derived from the bodily actions that **Agt** is intended to perform upon the intended referent.[9]

Finally, although not stated in this definition, the means by which the act is performed is some function mapping **D** to some plan or procedure that, when executed by **Agt**, enables **Agt** to discover the **X** that is the referent of **D**.

Even with this imprecise understanding of referent identification, it is apparent that not all noun phrases used in task-oriented conversations (even with the perceptual access conditions satisfied) are uttered with the intention that their referents be identified. For example, in dialogues with an information booth clerk in a train station (Allen 1979, Horrigan 1977), patrons uttering "the 3:15 to Montreal?" are not intending the clerk to pick out the train. Instead, as part of their plan for boarding a train, patrons are intending the clerk to supply them with a co-referring noun phrase that will allow them to identify the train. The attributive use of definite noun phrases (Donnellan 1960) is another case in which the speaker has no intention that the hearer identify a referent. Other non-anaphoric uses of noun phrases include labeling an object, correcting a referential miscommunication, getting the speaker to wait while the speaker identifies the referent, etc.[10]

3.2 Comparisons with computational linguistics approaches to reference

Computational linguistics research has usually been concerned with co-reference – the relationship of words and symbols to other words and symbols. Typically, referents of descriptions are determined by intersecting the extensions of the predicates in the description, subject to the quantificational constraints imposed by the determiner. Although perhaps adequate for interfacing with databases, this approach presupposes that the extensions can be computed from information currently in the database. However, in interpreting and generating discourse about some physical task, the system may have to form a plan that it or its user perform physical actions to determine the extensions of the predicates.

Five approaches are most closely related to ours. First, Winograd's SHRDLU (1972) attempted to simulate true reference with co-reference.[11] SHRDLU had a PLANNER function, THFIND, that could find objects in the database satisfying THGOAL statements as a simulation of finding

[8] Actually, Moore characterizes **RESULT** as taking an event and a formula as arguments. In his framework, an agent's doing an action denotes an event. However, this difference is not critical for what follows.

[9] The connection with the contextually relevant actions is a matter of inference (see section 6).

[10] For other discussion of speakers' goals in uttering noun phrases (see Sidner (1983) and Wilkes-Gibbs, unpublished ms).

[11] On the other hand, one might argue that SHRDLU engaged in true reference because the discourse was about non-existing blocks "contained" within the system. To pursue the truth of the matter would take us too far afield.

blocks in the real world. THFIND was included in the semantic representation of definite NPs, and in the representation of indefinite NPs when those NPs were embedded in an action verb. However, THFIND is not attributed as a user goal, nor is it reasoned about (other than to maintain a distinction between definite and indefinite NPs). Furthermore, it is not treated in the same way as acts such as PICKUP, whose execution is marked specially so that the system can later answer "why" questions.

Second, Allen's (1979) system used an IDENTIFY state in the control part of the plan-recognition mechanism. Again, for this system, identification meant to find something in the database satisfying the requisite predicates. However, the IDENTIFY action itself was not part of the plan being recognized. The system did not reason about when IDENTIFY should be done (it always tried to IDENTIFY referents), nor did it attribute IDENTIFY to be part of its user's plan.

The TDUS system (Robinson et al. 1980) engaged in a dialogue about the assembly of an air compressor that, it was understood, was being assembled by an apprentice. Thus, the referents of the system's noun phrases were perceptually accessible to the hearer. The system was primarily oriented towards utterance interpretation, but it did generate responses to questions. In doing so, the system was in the same circumstances as the experts in the present study. However, because it was assumed that the extensions of all of the system's descriptors were already known to the hearer, the system did not reason that it should choose particular referring expressions so that the hearer could pick out their referents. Instead, the choice of referring expressions was constrained by uniqueness and focus (Grosz 1977), constraints that are not considered here but are clearly necessary. Although TDUS employed the concept of locating an object in its representation of successful task performance, this concept did not play a role in choosing referring expressions unless the system was asked a question about an object's location.

Appelt's KAMP system (1981) generalized TDUS to plan referring actions as part of the planning of illocutionary acts. However, KAMP would only include descriptors in a referring expression for which it was already mutually believed that the hearer knew the referent. Thus, it could not generate referring expressions to new objects for the hearer to pick out. Furthermore, as argued earlier, the concept of "knowing what the referent is", which was central to KAMP's planning of referring phrases, is too strong to be an accurate representation of referent identification.

Finally, the HAM-ANS question-answering system (Hoeppner, Morik, and Marburger 1984) generates descriptions of objects in a hotel room from visually derived information, assuming the user's visual search processes are identical with its own. In another application, the system answers questions about traffic flow based on visual data. In its tying reference to perception, the HAM-ANS system has some of the flexibility that I am advocating. However, as with the others, it does not reason about identification as an action that the speaker intends it to do. In this paper, I argue why such reasoning is needed.

3.3 Summary

In summary, I am suggesting that referent identification be an action that the hearer infers to be part of the speaker's plan, and that speakers plan for hearers to perform. To ensure that hearers can do so, speakers employ their knowledge of the hearer's perceptual abilities, and choose descriptions that will make use of those abilities. The ability to reason about the referent identification act will allow the hearer to infer the intentions behind many utterances that secure reference separately from predication, and do so indirectly. With this concept in mind, we can proceed to examine its use in discourse.

4 The Study

Twenty-five subjects ("experts") each instructed a randomly chosen "apprentice" in assembling a toy water pump, following Grosz's (1977) and Chapanis et al.'s (1972) task-oriented dialogue paradigm.[12]

Subjects were paid volunteer students from the University of Illinois, all of whom were familiar with CRT terminals. Five "dialogues" took place in each of the following modalities: face-to-face, by telephone, keyboard ("linked" CRTs), (noninteractive) audiotape, and (non-interactive) written. In all modes, the apprentices were videotaped as they followed the experts' instructions.

Face-to-face and written modalities are the ones usually compared in oral/written discussions. However, they differ along many dimensions (Rubin 1980). Pairwise comparisons of the modalities in this study can determine the effects of mutual vision, interaction, and the use of voice or print. Telephone and keyboard dialogues are analyzed first because our conclusions would indicate the effects of having a voice channel, and moreover would have implications for the design of speech understanding and production systems. These modalities take on intermediate values in Rubin's dimensional space: the conversants share the same time frame, can interact, cannot see each other, and are conversing about objects mutually known to be physically present to one of them.

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

[12] An exploded parts diagram of the pump can be found in Appendix A.

the pump as often as desired, and then instructing a research assistant. This practice session took place face to face. Experts knew that 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. Experts were not informed of the modality in which they would communicate until the next day.[13]

Apprentices were told the purpose of the experiment was to analyze the communicating of a set of instructions in different modalities. They were not initially informed that they were engaged in an assembly task.

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 through a standard telephone and apprentices communicated through a speaker-phone. This device did not need to be held and allowed simultaneous two-way communication. Distortion of the expert's voice was apparent, but not measured.

Subjects in "keyboard" 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.

4.1 Sample transcripts

The following are representative samples of transcripts in the two modalities.

A TELEPHONE DIALOGUE 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: OK. Now. now, the smallest of the red pieces?

J: OK"

A KEYBOARD DIALOGUE FRAGMENT

B: "fit the blue cap over the tub end

N: done

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

N: ok

•

•

B: right Put the 1/4 inch long 'post' into the loosely fitting hole...

N: i don't understand what you mean

B: the red piece, with the four tiny projections?

N: OK

•

•

B: place it loosely into the hole on the side of the large tube...

N: done

B: very good. See the clear elbow tube?

N: yes

B: place the large end over that same place.

N: ready

B: take the clear dome and attach it to the end of the elbow joint..."

4.2 Method of analysis

Discourses are analyzed for many reasons, with a corresponding variety of methods. Some analyses of discourse strive to explain what the text itself meant. Recent work on discourse pragmatics emphasizes the need to explain what the speaker meant in producing the utterances, i.e., what were the speaker's intentions? To build dialogue systems, we need to devise first theories and then algorithms for deriving what the speaker meant as a function of what was said and of contextual factors. The logical first step toward such a formalization is to establish reliable methods for isolating the words, context, and speaker intent. Each of these aspects of the discourse is considered below.

[13] The instructions given to the expert about the experiment and the assembly task are given in Appendix A. Burke (1982) reports that the order of the instructions, and the descriptions of the pieces, influenced the order and vocabulary of the expert's subsequent instructions.

First, the typewritten transcript of a verbal interaction provides reasonably accurate data on what was said, provided one's goal is not to study prosody. Second, contextual factors can be modelled in a setting in which the objects, communication task, and modality have been selected by the experimenter. The conversants' knowledge of the domain is somewhat constrained by the experimental setup and the initial instructions. This semi-controlled environment can enable the experimenter to model the participants' initial experiment-induced beliefs, intentions, and expectations, which constitute our model of the cognitive effects of context.

Finally, as the conversation progresses, one needs interpretations of what each speaker meant, stated in terms of further attributions of beliefs and intentions. Standard empirical methods should be used to minimize experimenter bias in making such attributions. In particular, the theorist must be careful not to be the source of belief/intent attributions, for if given the leeway, he will undoubtedly find what he is looking for. To avoid this problem, I trained two people to employ a vocabulary for describing intentions in discourse, the so-called "illocutionary acts" (or, loosely, "speech acts") (Austin 1962, Searle 1969). That is, the discourse analysts "code" the speaker's intentions in making an utterance by assigning illocutionary act labels to utterances (or groups of them). Fortunately, the illocutionary act vocabulary is the natural one in our common-sense psychology for making such attributions. However, unlike most theories of illocutionary acts, I do not claim that the conversants themselves attempt to determine what illocutionary acts were performed, although they might be able to do so if requested.[14] The illocutionary act interpretations are therefore our interpretations, as coders and as theorists.

The data that need to be compared and explained are these illocutionary act codings. As mentioned earlier, a number of researchers have attempted similar analyses, but are content with solely identifying regularities in their discourses. A preferable analysis would derive regularities from more basic principles. The method employed here for formulating such derivations includes the following components:

- A logic of beliefs, mutual beliefs, and goals.
- A specification of the goals achieved by utterances of various forms (e.g., a yes/no question is an attempt to get the hearer to inform the speaker whether or not the proposition in question is true).
- A formal theory of rational, intentional action that specifies how an agent's actions are determined by both his goals and his knowledge of the effects of, preconditions for, and means of accomplishing various action types.

The aim of a *competence theory* of communication based on plans is to specify the set of possible plans underlying the appropriate use of various illocutionary acts. In applying such a theory to the analysis of discourse, plans are used to connect an utterance's form and content with the observers' illocutionary act coding, which is our best approximation to the speaker's intent. It is important to remember that these intentions may not be identical to those conveyed by the literal utterance. The plans make use of a formalization of the experimental task, the modality and the prior discourse, expressed in terms of the participants' mutual beliefs, goals, expectations, and possibilities for action. Thus, the theory captures, albeit in an indirect way, the dependence of the discourse structure on the experimental task and communication modality.

In addition, a *performance model* would include algorithms for forming and recognizing plans of action to derive the observer's intent codings. Although such models have been built (Allen 1979, Brachman et al. 1979), I do not discuss them further here.

In summary, the discourse analysis methodology is as follows:

- Train coders to identify various illocutionary acts (IAs).
- Compare the distribution of IAs across modalities.
- Independently, characterize those IA types in terms of plans.
- Formally derive the IA codings as a rational strategy of action, given attributions of the participants' beliefs, goals, and expectations at the point in the discourse in which the IAs occurred.

When our work is complete, we will have analyses of the differences in achievement of the same overarching set of goals (the assembly task) as a function of modality.

4.2.1 Coding the transcripts

The first stage of discourse analysis involves the coding of the communicator's intent in making various utterances. Following the experiences of Sinclair and Coulthard (1975), Dore et al. (1978), and Mann, Carlisle, Moore, and Levin (1977), a coding scheme was developed and two people were trained in its use. The coders relied on written transcripts, audiotapes, and on videotapes.

The scheme, which was tested and revised on pilot data until reliability was attained, included a set of approximately eight illocutionary 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). Appendix B lists the propositions and operators for the physical actions. For example, putting two hollow, pipe-like pieces together was termed

[14] See Cohen and Levesque (1980, in preparation) for a plan-based theory of communication that does not require the recognition of illocutionary acts.

CONNECTing; putting a part with a protrusion into a part with a hole was termed MESHing. The operators for physical actions often served as the propositional content of the communicative acts.

The following illocutionary act categories were coded:

Communicative Act
Example

Request(Assembly Action)
"put that on the hole"

Request(Orientation Action)
"the other way around"
"the top is the bottom"

Request(Pick-up)
"take the blue base"

Request(Identify-Referent)
"there is a little yellow piece of rubber"

Request(Informif([relation]))
"and you've got the base on it?"

Request(Informif(Identified-referent))
"got it?"
"the little red plug?"

Request(Achieve([relation]))
"and the purpose of that is to cover up that hole....
[relation] = (Cover V2 Hole(TB)))]

Label
"that's a plunger"

As discussed earlier, the action of referent identification is labelled IDENTIFY-REFERENT, and the state of affairs resulting from it is termed IDENTIFIED-REFERENT. Communicating that the speaker wants the hearer to do something is termed REQUESTing. Yes/no questions are REQUESTs to get a hearer to perform an INFORMIF action, i.e., to tell the speaker whether or not some proposition holds. One subcase of this is to tell the hearer whether or not a referent for a description has been identified. Finally, speakers often request that hearers make a relation true, without specifying an action that would do so. This is captured by the REQUEST to ACHIEVE [relation] coding.

Regarding referent identification, the coders were asked to state which utterances, or groups of utterances, constituted either an *explicit request* by the speaker that the hearer identify the referent of a noun phrase or a question about whether or not the hearer had done so. The coders were instructed not to consider whether or not an utterance was an indirect request to pick something up (but see section 6.4.1). Furthermore, they were told not to consider noun phrases in assembly requests as identification requests unless identification was somehow "explicitly marked".[15] Because agreement about the intent behind utterance parts was not obtainable, I cannot assert, on the basis of empirical evidence alone, that noun phrases embedded in imperatives are requests to identify the referents. Instead, the speaker's intent behind whole utterances (though not necessarily complete sentences) was coded.[16]

4.2.2 Mechanics of coding

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. The coders were instructed to ignore false starts, even though a false start may communicate information.

Although our goals did not include a precise analysis of how prosody reflects speaker-intent and meaning, some decisions about how to translate prosody into orthographic form, which undoubtedly influence subsequent discourse analyses, were made by the transcriber of the audiotapes. To minimize inconsistencies in transcription, all transcriptions were checked by a second party. Moreover, 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 to yield a base set of codable utterance parts. The coders could later bracket utterances differently if necessary.

For one third of the transcripts, interrater reliabilities were calculated within each mode, for each category. The measure consisted of twice the number of agreements divided by the number of times that category was coded (cf. Mann, Carlisle, Moore, and Levin 1977). Reliabilities were high (above 88%). Because each disagreement counted twice (against both categories that were coded), agreements also counted twice.

4.2.3 Coding the sample dialogue fragments

The previous fragments are coded below to indicate some of the complexities of the data as well as the scoring scheme. A number of shortcuts have been taken for expository purposes. First, if an act is stated as

[15] The above Telephone dialogue fragment contains one such intonationally marked noun phrase.

[16] For a formal analysis that does make such a claim, see section 8.4 and Cohen (1984).

[17] The action-effect relation holding between the various propositions and assembly actions can be readily inferred from Appendix B.

COMPLETE, then the proposition stated as the effect of that act holds.[17] Second, some of the arguments to the embedded propositions have not been presented when those arguments are not problematic. Third, as argued above, the second argument of IDENTIFY-REFERENT should be a description in some appropriate logical form representing the meaning of the speaker's noun phrase. However, because it was too difficult to get coders to determine logical forms for the noun phrases, they instead coded only the canonical names of the referents as arguments. Finally, the elapsed time between utterances is not shown here, but is available from the videotapes.

The codings of S's first turn indicate an attempt to achieve more than one intention in one utterance. Specifically, the form "there's a ...", is a typical way to perform a request to identify something satisfying the description (the "..."). In this case, the speaker said "thing", and labelled that thing a plunger. Whereas the labelling act may be finished, the request for referent identification apparently is not, and is continued over a number of utterances.

The other "bracketed" turn is an example of a speaker's prosodically achieving multiple goals at once. Here, the use of rising intonation in the middle of an imperative is used to check whether the hearer knows what the speaker is talking about. The pragmatics of this discourse situation led to the coding of "knowing what the speaker is talking about" as a request to physically identify a referent. Finally, notice the subsequent use of a questioned noun phrase fragment to perform the same act. The use of fragments will be discussed further below.

The coding of the Keyboard utterances is more straightforward. 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 (12). Finally, B performs separate identification requests, as in (7) and (8).

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

In (5)-(7), we observe clarification dialogue about a noun phrase. Immediately after an apparent breakdown at (3), B "downshifts" to questioning the achievement of his first subgoal, identifying the red piece. Once that is corrected, B stays in "low gear", explicitly ensuring success of his reference, in (8), before requesting an assembly action in (9). After that success, he "upshifts" to "second gear" – with requests to pick-up and assemble in (13). After being successful yet again, B "upshifts" to "high gear", using direct assembly requests, for the rest of the dialogue (seven more requests).

What could explain this conversation pattern? A common sense analysis of the plan for assembling would indicate 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 infer the rest of the plan. By requesting the sequence take-and-assemble ("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 identifying the referents of the speaker's descriptions. In summary, the strategy shift to "low gear" occurs after a referential miscommunication because it affords a more precise monitoring of the listener's achievement of the speaker's goals. The question to be asked is how, if at all, the use of identification requests differs across modes of communication.

5 Results

5.1 Analysis 1: Distribution of requests

Because most of each dialogue consists 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 1 displays the findings.

Identification requests, which include questions about whether a referent has been identified, are much more frequent in Telephone dialogues than in Keyboard conversations. In fact, they constitute the largest category of requests in the former. Because orientation requests, pick-up requests, and other requests are often issued to clarify or follow up on a previous request, it is not surprising that they would increase in number (though not as a percentage) with the increase in the use of identification requests. Furthermore, it is sensible that there are about the same number of requests for assembly actions (and hence half the percentage) in each mode because the same "assembly work" is accomplished. Therefore, identification requests seem to be the primary request differentiating the two modalities.[18] Notice also that Chapanis et al.'s finding of twice as many words used in spoken over written modes holds true when we consider the number of requests rather than just words.

[18] The only cases of unreliable coding resulted from attempts to code identification requests when they were not obviously separate utterances. Most of these unreliable cases were found in the Keyboard dialogues, but were included in the totals for that modality. Thus, using a strict criterion of counting only reliable examples of identification requests, the differences between the two modes are even stronger than shown in the table.

	UTTERANCE	CODING

S:	"OK. Take that.
	Now there's a thing
	called a plunger.
	It has a red handle on it,
	a greenbottom, and it's got a blue lid.
		REQUEST(PICK-UP)
		LABEL(PLUNGER)
		REQUEST(IDENTIFY-REFERENT(PLUNGER))

J:	OK	INFORM(COMPLETE (IDENTIFY-REFERENT(PLUNGER)))
	•	
S:	OK now, the small blue cap we talked about before?	REQUEST(INFORMIF (IDENTIFIED-REFERENT (TUBE-CAP)))
J:	Yeah	INFORM(COMPLETE (IDENTIFY-REFERENT(TUBE-CAP)))
S:	Put that over the hole on the side of that tube –	REQUEST(COVER) REQUEST(INFORMIF (IDENTIFIED-REFERENT OUTLET1))
J:	Yeah	INFORM(COMPLETE (IDENTIFY-REFERENT (MAIN-TUBE)))
S:	– that is nearest to the top, or nearest to the red handle.	
J:	OK	INFORM(COMPLETE (COVER))
	• •	
S:	OK. now. now, the smallest of the red pieces?	REQUEST(INFORMIF (IDENTIFIED-REFERENT (PLUG)))
J:	OK	INFORM(COMPLETE (IDENTIFY-REFERENT(PLUG)))

Figure 2. *The telephone fragment coded.*

5.2 Analysis 2: First-time identifications

Although frequency data are important for computational linguistics, they supply only a coarse description of discourse phenomena and the dialogue itself. For assessing the importance of a phenomenon, it is important to discover the context in which it occurs. In our case, the question arises whether the frequent use of identification requests is a function of the modality itself or a function of the dynamics of the discourse (or both). For example, identification requests might arise primarily after referential miscommunication, as in the above Keyboard dialogue. If they did, one might argue that people would speak more carefully to machines than they do to people, leading to less miscommunication and therefore a smaller frequency of identification requests than we found. Consequently, the argument that identification requests are important for computational linguistics research would be weakened. However, Fertig (unpublished) found no significant differences in the frequency of miscommunication across modes. The observed modality differences in the use of identification requests are therefore due to other causes.

To determine if the frequent use of identification requests holds across subjects within modes, a second analysis of the utterance codings was undertaken that was limited to "first-time" identifications – that is, how objects were first introduced to hearers. Each time a novice first identified a piece in response to a communicative act, that act was noted. Furthermore, that act was counted only if it was not preceded by another mentioning the same part prior to the novice's identification attempt. This analysis therefore examines only requests used to make *first effective* reference to an object. Table 2 indicates the results for each subject in Telephone and Keyboard modes.

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,

UTTERANCE	CODING
(1)B: “fit the blue cap over the tub end	REQUEST(COVER(TUBE-CAP MAIN-TUBE))
(2)N: done	INFORM(COMPLETE(COVER))
(3)B: put the little black ring into the large blue cap with the hole in it...	REQUEST(PUT-INTO (O-RING TUBE-BASE))
(4)N: ok	INFORM(COMPLETE(PUT-INTO))
• •	
(5)B: right put the 1/4 inch long ‘post’ into the loosely fitting hole...	REQUEST(MESH (VALVE3 OUTLET2))
(6)N: I don’t understand what you mean	
(7)B: the red piece, with the four tiny projections?	REQUEST(INFORMIF (IDENTIFIED-REFERENT (VALVE3)))
• •	
(8)B: very good. See the clear elbow tube?	REQUEST(INFORMIF (IDENTIFIED-REFERENT (SPOUT)))
(9)N: Yes	INFORM(IDENTIFIED-REFERENT (SPOUT))
(10)B: Place the large end over that same place.	REQUEST(CONNECT (SPOUT OUTLET2))
(11)N: ready	INFORM(COMPLETE (CONNECT SPOUT OUTLET2))

(12)B: take the clear dome and attach it to the end of the elbow joint...	REQUEST(PICK-UP(AIR-CH))
	REQUEST(CONNECT(AIR-CH SPOUT))

Figure 3. *The keyboard dialogue fragment coded.*

four of five experts were habitual users of identification requests to get the apprentice to find a piece. The remaining subject used the strategy of first requesting the apprentice to pick up a part, and then requesting that it be attached to the pump. In Keyboard mode, no experts were habitual users of identification requests. However, three experts were habitual users of assembly requests in getting apprentices to identify objects.

To show a “modality effect” in the making of first effective reference, the number of habitual users of each request type in each mode was subjected to Fischer's exact probability test. This calculates the probability that differences in the number of habitual and nonusers of a particular reference strategy in different modes could have happened by chance. Even with five subjects per mode, differences in the use of identification requests across modes were significant ($p = 0.023$), indicating that Telephone conversation per se differs from Keyboard conversation in the ways in which a speaker will first get a hearer to identify an object.

In summary, it has been shown that Telephone and Keyboard modes differ primarily in the use of explicit identification requests. These requests do not simply occur after referential miscommunication (as they do in Keyboard), but are used to first introduce objects. The experts then often question the apprentices about successful completion of the identification act (just as they do assembly acts). Experts using keyboards do not attempt to achieve referential goals explicitly. Instead, referential goals are subsumed in assembly requests. Voice communication is thus “finer-grained” than keyboard communication.

5.3 Comparison with other studies

These results are similar to observations by Ochs and colleagues (Ochs 1979; Ochs, Schieffelin, and Pratt 1979). Using evidence from transcripts, they point out that caretaker-child and child-child discourse often consists of “sequential” constructions – with separate utterances for securing reference and for predicating. Ochs et al. suggest that the presence of sequential

Table 1. *Distribution of requests (percent).*

Type of Request	Telephone (n=288)	Keyboard (n=134)
Assembly	25	51
Orient	9	8
Other	15	13
Pick-up	16	17
Identification	35	10

constructions is tied to the possibility for preplanning an utterance. Relatively unplanned discourse, it is claimed, relies on the pragmatic context to express propositions, where planned discourse would use syntactic means. Unplanned discourse results when speakers are concentrating on a task or when the expression of a concept is particularly difficult. The present study upholds Och et al.'s claim for Telephone and Keyboard communication, but does not do so for the Written condition, in which many identification requests occur as separate steps (Tierney et al. 1983). Furthermore, Ochs et al.'s claim does not account for the use of identification requests in Keyboard modality after prior referential miscommunication (see section 4.1 for a sample conversation), indicating that sequential constructions can result from (what they term) planned as well as unplanned discourse.

Clark and Wilkes-Gibbs (unpublished) analyze referential communication data similar to ours. Their concern is to show that referring is a collaborative process, one that proceeds by a speaker's proposing a referring expression, and a hearer's accepting or rejecting it as adequate for identifying the referent. Among the speaker's strategies for securing reference, they note elaborations (which I called "supplements"), trial proposals (Question-requests for identification), partial proposals, and others. Unlike what has occasionally been assumed in the referential communication literature, speakers are not regarded as trying to produce, in one turn, an effective referring expression that is minimally long. Instead, they claim speakers attempt to minimize the collaborative effort of both parties. The discourses they analyzed overlap significantly in structure with those found here. The present account differs in that I give a formal analysis of the act of referring as an illocutionary act – an account that allows for indirect performance.

It is difficult to compare the present results with those of other studies. Chapanis et al.'s (1977) observation that voice modes are faster and wordier than keyword modes certainly holds here. However, their transcripts cannot easily be used to verify the present findings because, 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.

Stoll et. al.'s (1976) lexical analysis of the Chapanis data indicates that in verbal modalities subjects produce many more pronouns and (what they term) "function words", which include articles, prepositions, modals, existential "there", etc. In the present study, there were at least two requests used for each assembly step in Telephone mode. Each pair of requests (identification requests followed by assembly requests, or requests to pick up followed by assembly requests) involved at least one common object being manipulated. In the assembly request, the speakers frequently referred to that object with a pronoun. Thus, because of the pragmatically fine-grained nature of Telephone mode, there are many more pronouns to to resolve. I suspect the same kind of analysis can be applied to Stoll et. al.'s "function word" category, although that category is so diverse that generalizations may be harder to find. However, it is clear that in the present data, the use of "existential there" sentences, by far the largest class of identification requests, is only a Telephone strategy.

Modality differences in Grosz' (1977) study cannot be directly compared for the identification phenomena because the core dialogues that were analyzed in depth each employed both spoken and keyboard modalities. However, the present results would predict that indirect identification requests would not appear because the expert, who did most of the communicating, used a keyboard.

Finally, Thompson's (1980) extensive tabulation of utterance forms in a multiple modality comparison overlaps the analysis in this paper at the level of syntax. Both Thompson's and the present study are primarily concerned with extending the usability of current systems by identifying phenomena that people use, but that would be problematic for computers. However, the 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. This study's concern stems from the observation that differences in utterance function will influence the processing of utterances with the same

Table 2. *Communicative Acts Making First Effective Reference to Each of 12 Pump Pieces.*

	TELEPHONE REQUESTS			KEYBOARD REQUESTS		
SUBJECT	IDENT	PICK-UP	ASSEMBLY	IDENT	PICK-UP	ASSEMBLY
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

form. The remainder of this paper explores issues of inferring utterance function partly from utterance form.

6 Analysis of Utterance Forms: Identification Requests

Thus far, explicit identification requests have been shown to be pervasive in Telephone mode. One might expect that, in analogous circumstances (i.e., with analogous goals and perceptual capabilities), a robot might be confronted with many of these acts. Computational linguistics research then must discover algorithms for determining an appropriate analysis and response, in part as a function of utterance form. To see just which forms are used for the task, utterances classified as identification requests in Telephone mode were tabulated. The full listing of identification requests can be found in Appendix C.

Table 3 presents a classification of these utterances, along with an example of each class. The utterance forms are divided into four major groups, based on their similarities in either syntactic or logical form. The first group consists of utterances whose logical form is an existential proposition (usually determined by the presence of an indefinite noun phrase). The second category includes those utterances that mention perceptual actions (e.g., "look") and perceptual effects (such as "see"), but are not syntactically imperatives. The third class contains identification requests performed with utterance fragments, usually noun or prepositional phrases (PPs). The concept of fragment is not solely a syntactic classification – hearers can be requested prosodically to respond to NPs and PPs that are embedded in full sentences. Other categories of utterance forms for identification requests include what are "nearly direct" requests (i.e., imperatives and utterances that explicitly mention searching for an object) and what are termed "Let's requests", which explicitly change the focus of attention to an object satisfying the description. Finally, one class of utterances, accounting for 11% of identification requests and 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. Typically, such utterances were coded as questions and also as requests for identification.

The important thing to notice in Table 3 is that in Telephone mode identification requests were *almost never* performed directly. No speaker used direct forms, e.g., "Find the rubber ring shaped like an O", which occurred frequently in the Written modality (Tierney et al. 1983). However, the use of indirection is selective – Telephone experts frequently use direct imperatives to perform assembly requests. Moreover, many speakers adopted a consistent style. For example, all the "nearly direct requests" came from one speaker, and another almost uniformly used the "there's a NP" strategy.

Because explicit identification requests come in many syntactic forms, each of which has a literal interpretation that is not an identification request, the hearer needs some method for deciding what the speaker's intention(s) are. Ideally, such reasoning should be an application of more general reasoning about nonlinguistic actions. Of course, a suitable "compiling" strategy can specialize the general case for this task (cf. Brachman et al. 1979). Below, I sketch such a theory and apply it to the examples in Table 3.

6.1 A sketch of a plan-based theory of communication

The unifying theme of much current pragmatics research is that the coherence of dialogue is to be found in the interaction of the conversants' *plans*. That is, a speaker is regarded as planning his utterances to achieve his goals, which may involve influencing a hearer. On receiving an utterance, the hearer attempts to infer the speaker's goal(s), and to understand how the utterance furthers them. The hearer then adopts new goals (e.g., to respond to a request, to clarify the previous speaker's utterance or goal), and plans his own utterances to achieve those. A conversation ensues.

Table 3. *Identification requests in telephone mode.*

Group	Category [Example]	Per Cent of Request(Ident.)
A.	EXISTENTIAL PROPOSITIONS	
	1. THERE'S A NP(?) ["there's a black o-ring (?)"]	25%
	2.OBJ HAS PART(?) ["It's got a peg in it"	14%
	3. LISTENER HAS OBJ? ["Now you have two devices that are clear plastic“]	10%
	4. DESCRIPTION1 = DESCRIPTION2 ["The other one is a bubbled piece with a blue base on it with one spout"]	9%
B.	PERCEPTION-BASED	
	1. INFORM(IF ACT THEN EFFECT) ["If you look at the bottom you will see a project"]	2%
	2. QUESTION(EFFECT) ["If you look at the bottom you will see a projection"]	5%
	3. INFORM(EFFECT) ["you will see two blue tubes"]	2%
C.	FRAGMENTS	
	1. NP AND PP FRAGMENTS (?) ["The smallest of the red pieces?"]	12%
	2. PREPOSED OR INTERIOR PP (?) ["and in the bottom of the blue cap on the main tube (pause) there is a hole"]	6%
D.	NEARLY DIRECT REQUESTS	
	["Look at the bottom of the tube"]	1%
	["The next thing you're gonna look for is...."	1%
E.	LET'S REQUESTS	
	["Let's go to the little tiny blue cap"]	5%

Recent work of this type (Allen 1979, Appelt 1981, Bruce 1983, Bruce and Newman 1978, Bruce and Schmidt, Cohen 1978, Cohen and Levesque 1980, Cohen and Perrault 1979, Perrault and Allen 1980, Schmidt 1975, Sidner and Israel 1981) has resulted in formal and computational models of communication that have been applied in analyzing dialogues about tasks and stories. The general features of the models include: a simple theory of action, definitions of various physical and communicative actions, a set of inference rules for formulating and recognizing plans of action, a formalization of agents' beliefs, goals, and expectations, and a mapping of utterance forms to the "surface speech actions" speakers are performing in making those utterances.

For the purposes of this paper, *planning* is simplistically viewed as the process of finding an action (or a sequence of them) that will achieve the agent's goal(s) given what he believes to be the state of the world. Roughly speaking, to *recognize* an agent's plan in performing an action, observers deploy a theory of planning "in reverse" to connect the observed action with a chain of inferences of the form "agent did X in order to achieve Y, which would enable him to do Z", terminating in (what they take to be) a likely or expected goal of the agent (Allen 1979; Genesereth 1978; Schmidt, Sridharan, and Goodson 1979; Wilensky 1978). Such reasoning employs beliefs about the agent's beliefs, conditions that are likely to be true at the end of an action, other actions that are enabled by those conditions, and expected plans and goals of that agent.

6.2 Analyzing indirect identification requests

Perrault and Allen (1980) have proposed the following plan-recognition inferences:

- *Action-effect:* If the observer thinks the agent wants to do an action, the observer can posit that the agent wants that action done in order to achieve its typical effect.
- *Precondition-action:* If the agent is thought to want some proposition P to be true, and P is the precondition of an action known to the agent, consider that the agent's goal is to perform that action.
- *Body-action:* If the agent is thought to have a goal that is the means by which a "higher level" action is performed, then consider that the agent is attempting to perform that action.[19]
- *Knowif:* If the agent is thought to want to know whether or not some proposition P holds, then consider that the agent wants P to hold (or wants P to be false).

The plan-recognition process first classifies utterances by their mood into so-called "surface speech acts": declaratives become instances of the S-INFORM act type, imperatives become S-REQUESTS, and questions become S-REQUESTS to INFORMIF or INFORMREF (for Yes/No and Wh questions, respectively). These surface acts are the prototypical ways to perform the corresponding illocutionary acts REQUEST, INFORM, and REQUEST to INFORMIF.[20]

Grice (1957) has argued that a simple plan-recognition process, which an unseen observer might perform, cannot be the basis for communication. Rather, the hearer must infer and act on what the speaker wants him to *think* she wants. A plan-based theory proposes that such reasoning is invoked by applying the independently motivated plan-recognition process to the observed communicative actions. Hearers are, in effect, asking themselves "Why did the speaker say that?" Subsequent reasoning, termed here *intended plan recognition,* derives other goals that the hearer thinks she or he was *supposed* to infer. The process of attributing goals to an agent is terminated when an inference path merges with a mutually expected goal of that agent.

This plan-based approach has led to a first explanation of a large class of "indirect speech acts" – utterances whose surface form indicates one speaker intention, but for which additional intentions should be recognized. For example, although "Can you reach the hammer?" is literally a yes/no question, the speaker may have another goal – to get the hearer to pass the hammer to the speaker. The essential insight of the theory is that indirect speech act recognition is a by-product of the general process of recognizing someone's plans. If illocutionary act identification occurs immediately after the expansion of a surface act, then a *literal* interpretation has been found. If there are intervening intended plan-recognition inferences, then an *indirect* interpretation has been inferred.

6.3 Example

The following example is intended to give a brief introduction to the reasoning underlying indirection. Details of this process can be found in (Allen and Perrault 1980, Perrault and Allen 1980). The utterance to be interpreted is, *"Can you reach the hammer?"* The inferred propositions are indicated in **boldface;** commentary on the inference process is indented.

After parsing and semantic interpretation, the utterance is represented as a surface speech act:

**HBSW (S-REQUEST(S,H,INFORMIF(H,S,
CANDO(H,REACH(H,HAMMER)))))**

> That is, the *H*earer *B*elieves the *S*peaker *W*anted to perform (what appears to be) a yes-no question about his ability to reach the hammer.
>
> The effect of an S-REQUESTS to do action ACT is simply that the hearer believes the speaker wants the hearer to do ACT. Thus, applying the action-effect inference to the S-REQUESTS act, yields:

**HBSW(HBSW(KNOWIF(S,
CANDO(H,REACH(H,HAMMER)))))**

> The first level of "HBSW" comes from assuming the speaker's action was intentional. The second level is derived from the effect of the surface speech act. Now, intended plan recognition involves deriving new formulas of the form HBSW(HBSW(G)), i.e., the Hearer Believes the Speaker Wants him to think the speaker's goal is G. Any goals G' derived with the prefix HBSW(HBSW) are taken to have been communicated (in the Gricean sense). The truth of the preconditions to the actions that are inserted into the (intended) plan being recognized are evaluated with respect to the mutual beliefs of speaker and hearer.
>
> The two outer levels of "HBSW" are assumed to embed all the following formulas. In particular, applying the knowif inference at this level of embedding, the hearer realizes the speaker wants to knowif P (i.e., to know whether or not H can reach the hammer) because he wants him to be able to reach the hammer, yielding

[19] Note that this inference can adversely affect the combinatorics of the plan-recognition process.

[20] Levesque and I (in preparation) demonstrate how the intentions underlying these indirect speech act analyses can be derived using the surface speech acts and plan-recognition inferences, but not Perrault and Allen's illocutionary acts.

CANDO(H,REACH(H,HAMMER))

CANDO(H,ACT) is true when the preconditions of ACT are true, and the speaker wants the preconditions to be true because he wants the act done:

REACH(H,HAMMER)

S wants the act done for its effect:

HAVE(H,HAMMER)

S wants the effect because it enables another act:

PASS(H,S,HAMMER)

Finally, because the embedded **HBSW(PASS...)** is a way of performing a request to pass the hammer, the *body-action* inference yields:

HBSW (REQUEST(S,H,(PASS(H,S,HAMMER))))

Of course, the procedure could also have derived a request to reach the hammer. It does not do so because of the "level-of-inference" heuristic that pursues inference paths at the HBSW(HBSW) level of embedding before those with just the HBSW level of embedding. Applying the *action-effect* inference to the REQUEST action, and collapsing a few uninteresting inferences, H infers:

HBSW(PASS(H,S,HAMMER))

Then, H arrives at:

HBSW(HAVE(S,HAMMER))

Because the speaker's having the hammer was (assumed to be) an expected goal, the inference process stops.

To summarize, the plan-based theory views speech actions as planned actions that can be reasoned about in the same ways as physical actions. Indirect speech act interpretation involves linking the surface speech act characterizing each utterance with some expected goal through a chain of means-ends reasoning. When applied to a discourse, the plan-based theory attempts to mediate between utterance form and a (potentially changing) set of beliefs, goals, and expectations.

6.4 Applying the theory

Assume that syntactic and semantic components have already analyzed the utterances, resulting in a logical form for each sentence or complete constituent. Furthermore, assume that the apprentice has inferred the following expectations about the expert's goals: "For each piece making up the pump: The expert gets the apprentice to: identify the piece, pick it up, and perform some assembly action on it." Such expected goals can be used to terminate the process of plan recognition. Now, 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. For convenience, the definition of IDENTIFY-REFERENT, from section 3.1, is repeated below:

$$
\begin{aligned}
&\forall\ \mathbf{D\ Agt}\\
&\quad \exists\ \mathbf{X\ [PERCEPTUALLY\text{-}}\\
&\qquad \mathbf{ACCESSIBLE(X,\ Agt)\ \&}\\
&\qquad \mathbf{D(X)\ \&}\\
&\qquad \mathbf{IDENTIFIABLE(Agt,D)]}\\
&\qquad\quad \supset\\
&\quad \exists\ \mathbf{X\ [RESULT(Agt,}\\
&\qquad \mathbf{IDENTIFY\text{-}REFERENT(D),}\\
&\qquad \mathbf{IDENTIFIED\text{-}REFERENT}\\
&\qquad\quad \mathbf{(Agt,\ D,\ X)]}
\end{aligned}
$$

Figure 4. *The act of referent identification.*

6.4.1 Existential propositions

The utterances in Class A can then be analyzed as requests for IDENTIFY-REFERENT by applying plan recognition to the definition of the surface speech acts. Class A includes all declarative utterances whose logical form is an existential proposition (∃ X P(X)), which includes utterances of the form "there is a ....", "you have a ....", and "[object] has a .... [part]". These utterances would appear literally to be informative. However, they can be interpreted as requests that the hearer IDENTIFY-REFERENT of the description "the P" by reasoning that a speaker's wanting a hearer to believe that a precondition to an action (IDENTIFY-REFERENT) is true can communicate (in the Gricean way) that the speaker wants that action to be performed, provided the act's effect is mutually known to be an expected goal of the speaker's.

For example, "it's got a peg in it" is represented as ∃ x [INSIDE(x, PLUNGER) & PEG(x)]. Informing a hearer of this proposition yields sufficient conditions for inferring IDENTIFY-REFERENT (APPRENTICE, [INSIDE(x, PLUNGER) & PEG(x)]), whose effect unifies with a mutually expected goal. Thus, the existing indirection machinery can handle these cases.[21]

Yes/no questions can be recognized as requests for identification by the means of the same plan-recognition process. There are two cases. In one case, the hearer is modelled as inferring that the speaker wants the proposition in question to be true because it enables some action (i.e., IDENTIFY-REFERENT) to be performed that will yield a desired effect (IDENTIFIED-REFERENT). The inference takes hold because it is mutually believed that the speaker already knows the answer to his question. For example, in saying, "There's a little blue cap?" it is shared knowledge that the speaker already knows such a cap is in front of the hearer. Therefore, finding out whether there is such a cap could not be the speaker's goal.[22] In the second case, the hearer wants to know whether some proposition is true because he wants it to be true (or false), and wants the hearer to make it true (or false). In this way, questioning IDENTIFIED-REFERENT can be taken as a request for IDENTIFY-REFERENT.[23,24]

6.4.2 Perception-based utterances

Plan recognition can also suggest how Class B utterances all convey requests for referent identification. In this class, ACT = LOOK-AT, EFFECT = AGENT SEE X. Because LOOK-AT is one of the constituent acts of IDENTIFY-REFERENT, Perrault and Allen's "body-action" inference, given a formalization of perceptual actions and their relation to searching, should make the necessary connection – the speaker wanted the hearer to LOOK-AT something as part of his IDENTIFY-REFERENT act. Specifically, Case 1 ["If you look, you will see..." (If you don't, you won't)] again appears to be an informative utterance about a conditional. The speaker's intent that the hearer actually look *for* something is derived by an inference saying that if a speaker communicates that an act will yield some *desired* effect, then one can infer that the speaker wants that act performed to achieve that effect. Case 2 ("Do you see X") is again indirect because the hearer can truthfully answer "No" if he is looking out the window. Again the appropriate intention – that the hearer look for X – can be inferred by noticing that the speaker's questioning whether the desired effect of an act holds conveys the sense that the act itself is desired (e.g., "Is the garbage out?"). Case 3 ("You will see X") is similar to Case 1, except that the relationship between the desired effect and the action yielding that effect is presumed.

6.4.3 Fragments

Group C utterances constitute the class of fragments classified as requests for identification. Case 1 includes NP fragments, often 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.

Allen and Perrault's (1980) method of handling fragments involves unifying (in the technical sense of the term) the effects of the possible surface speech acts corresponding to the fragment with expected goals of the speaker. The result is a more fully specified goal that can, perhaps, be acted on. For questioned singular definite noun phrases (NP), their model proposes two possible incomplete surface speech acts:

S-REQUEST(S,H,INFORMIF(H,S, ϕ (ιxNPx))), and
S-REQUEST(S,H,INFORMREF(H,S, ιy[ψ (ιxNPx) = y])).

For example, on hearing "The Windsor train?" the system would initially take the speaker either to be asking a yes/no question about *some* property ϕ of the referent of "the Windsor train", or to be asking to know that the referent of the value of applying *some* function ψ to the referent of "the Windsor train". The respective effects of these surface acts would involve the hearer's thinking the speaker's goal is to know whether or not that predicate ϕ holds, or to know the identity of the referent of ιy[ψ (ιxNPx) = y])). The appropriate predicates (ϕ) and functions (ψ) need to be supplied from domain-dependent expectations. In the former case, the predicate ϕ might be (LAMBDA y (LEAVES-AT(y,4:40))), whereas in the latter case, the function ψ might be (LAMBDA y (GATE(y))).

Consider a questioned noun phrase, and just the yes/no question interpretation. The next step, according to the theory, is to infer that (it is shared knowledge) that the speaker wants (the hearer to think) that the speaker's goal is that ϕ (ιyNPy) be true. The properties ϕ needed to "fill in" such fragments come from mutually expected goals of the expert. The expected goal in question for this domain, which is (somehow) derived from the nature of the task as one of manual assembly, is (IDENTIFIED-REFERENT APPRENTICE D X), where D becomes bound to ιyNPy, X names the referent, and ϕ is IDENTIFIED-REFERENT.

[21] The formalism also predicts that, in this task, an indirect request to identify an object would also be an indirect request to pick up that object, because the line of inference is unambiguous and its end goal is mutually believed to be desired. If it were mutually believed that a possible end goal of the inference path (e.g., that the apprentice be holding the piece) was not desired (because, for instance, the piece was known to be hot), then the hearer would not infer he was supposed to pick up the piece. Although the coders were not asked to make this distinction, further analysis indicates the modality differences are stronger when pick-up requests are considered together with identification requests in Analysis 2. In Telephone mode, all 5 subjects were habitual users of either identification or pick-up requests, whereas no subjects were habitual users of either of those request types in Keyboard mode. Differences across modes are significant ($p = 0.004$).

[22] "Do you have homework to do?" is similarly identified as a request that you do your homework.

[23] A similar inference occurs in recognizing "Is the garbage out?" as a request to take out the garbage.

[24] Occasionally, speakers appear to ask "real" questions. For example, after requesting the hearer to identify a part, a speaker asked, "Do you see that?" I would have coded this as a true question, and not as an identification request, because the goal of an identification request (the speaker's communicating his intent that the hearer identify the piece) has already been achieved. The interrogative, then, is truly a question about whether the part has been identified. The coders, however, were not asked to make this distinction. The formalism makes predictions concerning which utterances were identification requests and which were only questions; since the codings did not reflect this difference, we have considered all questions to be identification requests. The number of cases of suspected true questions is small enough that the quantitative results are still valid.

From this, the apprentice infers that he should make (IDENTIFY-REFERENT APPRENTICE *i*yNPy X) true. 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 2, Group B, above). Thus, by my positing an IDENTIFY-REFERENT act, and by my assuming that the effect of this act is expected, the inferential machinery can derive the appropriate intention behind the use of a noun phrase fragment.

Notice that "fragment" is not a simple syntactic classification. In Case 2 ("In the green thing at the bottom [pause]" "Mmhm"), the speaker paralinguistically "calls for" a hearer response in the course of some *linguistically complete* utterance. Because of the nature of the task, Case 2 is coded as an identification request. A simple syntactic classification of fragments would not consider this request as fragmentary.[25]

This treatment of fragments is different from the usual one in computational linguistics, in which the fragment is matched into prior syntactic forms. Such an approach cannot work for fragments spoken without prior linguistic context, nor for fragmented adverbials whose coherence depends on a prior system action. Allen and Perrault's approach is an attempt to access the user's goal without reconstructing either a full syntactic or semantic analysis.

6.4.4 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 ("Look at the bottom of the tube") mentions "Look at", 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 ("The next thing you're gonna look for is...") is literally an informative utterance, though a request could be derived in one step. It is important that the frequency of these "nearest neighbors" is minimal (2%).

6.4.5 "Let's" requests

One speaker used "Let's" requests explicitly to shift the topic of conversation to one previously "closed" (Grosz 1977), and in the process to get the hearer to re-identify an object. Whereas other identification requests shift the topic as a by-product, this request seems literally to be a topic shift, with identification as a by-product.

6.5 Summary

The act of explicitly 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 accounts for 69% of the identification requests in the corpus [class A (56%), class C1 (12%), and class D2 (1%)]. Furthermore, plan recognition can infer the LOOK-AT action from the perception-based utterances [class B (9%)], but cannot yet connect LOOK-AT to IDENTIFY-REFERENT, because the means for performing IDENTIFY-REFERENT remains unspecified. Class C2 (preposed or interior prepositional phrases) requires a prosodic analysis to determine that a hearer response is called for and precisely what constituent is in question. With such an analysis, Perrault and Allen's fragment analysis can employ expectations to respond appropriately. Finally, Class E ("Let's" requests) is problematic for this theory (but see Grosz 1977).

7. Alternative Explanations Countered

I have argued that to interpret indirect identification requests, the apprentice needs to reason about the expert's intent. Alternative explanations could be envisioned that would place the burden of noun phrase interpretation entirely on the hearer. According to such accounts, the hearer would act on a noun phrase as he pleased, independent of the speaker's intended use of that noun phrase. A succession of such explanations is countered below.

1. *The hearer will identify the referent of every noun phrase.* Clearly, for both definite and indefinite noun phrases, this explanation is inadequate. Noun phrases such as "See the tapered red piece with the hole? That's the *nozzle*", and "We need *a valve* for that hole. It's the little yellow piece of rubber", supply functional descriptions or labels that *cannot* be identified at the time of utterance.[26] Rather, the speaker informs the hearer about some *future* function of that piece.

2. *The hearer identifies every description he can.* For example, perhaps a hearer will identify all things described with color or shape terms, or all noun phrases with existential presuppositions. However, this is again too simplistic. A hearer will undoubtedly not react to "there is a red cobra under that basket", by trying to identify the cobra because, in short, he doesn't want to.

3. *The hearer identifies every description he can identify and wants to identify.* That is, the hearer would

[25] Such examples of parallel achievement of communicative actions cannot be accounted for by any linguistic theory or computational linguistic mechanism of which I am aware. These cases have been included here because I believe that the theory should be extended to handle them by reasoning about parallel actions.

[26] The IDENTIFIABLE precondition was posited to prevent such noun phrases from being identified.

ignore what the speaker intends, and act on what he thinks is both feasible and desirable. To see that this cannot be the hearer's sole strategy, consider the following Telephone dialogue (and recall that the expert had a set of pieces in front of him, though the apprentice did not know this):

Expert: Now, take the big blue stopper that's laying around and take the black ring..."

Apprentice: [Searches and repeats more slowly] "the big blue stopper"

Expert:"Yeah, the big blue stopper [short pause] and black ring."

Although the apprentice could be said to be following the purported reference strategy in response to the expert's definite noun phrase, the same could not be said of the expert in responding to the *same NP*.[27] The expert is obviously not intended by the apprentice to identify the piece even though he could, because he has just requested the apprentice to do so, and because the expert's having a set of pieces was not mutually believed. The expert responds to (what he believes to be) the apprentice's purpose in repeating the NP, and not solely to his own desires and capabilities.

According to the above argument, the hearer cannot be completely egocentric in his interpretation of noun phrases, but considers the speaker's intentions with respect to that noun phrase.[28] One might still argue, however, that such consideration is quite simple; that a conversational "script" (Schank and Abelson 1977), involving the specification of the part the expert desires the apprentice to pick up and the assembly action to be performed on it, could handle the data. I argue that because the script notion of role already incorporates the expected goals of the parties who are playing each role, a script argument supports the position that noun phrase interpretation requires analysis of speaker intent.

4. *The apprentice fits the noun phrase into the script for his role in the experiment.* Scripts (Schank and Abelson 1977) contain expected sequences of actions, related in some "causal chain", in some stereotyped situation. According to Schank and Abelson, typical scripted situations might include birthday parties, restaurants, classrooms, etc. Scripts are parameterized by "slots" that define "roles" in the various actions and events. Essentially, a script's participants perform the specified actions, which, having been a successful pattern of interaction in the past, are already structured to achieve the goal(s) of the script. "Dialogue Games" Levin and Moore (1977) can be seen as scripts once utterances are viewed as communicative acts.

I claim that a script analysis of the discourses in the present experiment, if sufficiently detailed, supports the positions that (1) noun phrase interpretation requires an analysis of speaker intent, and (2) the speaker's intent is that the hearer perform an action of referent identification. The argument has three prongs: the contents of the purported script, the apprentice's inferring of that script, and the relationship of the script to the utterances.

First, as Schank and Abelson have pointed out, scripts are frozen plans. When two parties agree (perhaps tacitly) to play roles in a script, they have adopted their (respective) expected actions as part of their (respective) plans. This observation supports both of the above points. The contents of the script become expected goals, and the script contains *actions* that each participant is supposed to play. Any specification of the actions in the purported experimental script will contain the apprentice's performing IDENTIFY-REFERENT actions as a goal of the expert.

Second, the script needs to be inferred. The standard, well-worn, mundane activities (such as eating at a restaurant) that are claimed to be captured in scripts may not apply here. Apprentices may never have been in similar circumstances before (such as being in an experiment, or being instructed over a telephone), and thus may not have had a preexisting script. Furthermore, the apprentices were not told the script in the instructions nor by the expert (generally speaking, although there were a few exceptions). Thus, to the extent that a script is available to the participants, it would have been inferred.[29] Because scripts are frozen plans, to infer a script, apprentices would have engaged in a process of plan recognition in advance of engaging in the experimental task, or while the task was being achieved.

This process of inferring a script yields a set of expected actions. The plan-based theory of communication would require that the elements of a script be mutually expected and intended – i.e., that they would represent mutual beliefs about intended future actions. Such mutual expectations about each others' goals can terminate the process of plan recognition

[27] It is apparent that the two NP's are spoken with different intonation and timing. Prosodic aspects of an utterance are often regarded as signalling the speaker's attitude or intent toward what is being said. Thus, perhaps the speaker is prosodically signalling the "wait while I identify" intent. This conclusion supports the argument that speaker intent plays a role in processing descriptions.

[28] The "referential communication" literature considers the converse position – whether speakers are egocentric in producing noun phrases for a hearer (Asher 1979).

[29] No mechanisms have been proposed in the literature that can derive such expectations of the expert's goals from more basic information about the task, modality, genre, etc. At most, what has been proposed is the ability to use such expectations, independently of how they are derived.

applied to utterance interpretation. Thus, the inferring of a script that contains the expert's intent that the hearer identify the referents of descriptions is consistent with my proposal.

Although much of the determination of speaker intent has been precomputed, the utterance itself cannot be ignored. The results of this study show that speakers in the Telephone condition do not achieve their referential goals with direct requests, as they do in the Written condition (Tierney et al. 1983). The analysis of indirection used here (and similarly advocated by most speech act theorists) requires that speaker intent be recognized as a function, in part, of utterance form. Possible specifications of illocutionary force are derived from features of the utterance. Indirect speech acts are recognized through a chain of intended plan-recognition inferences (based on mutual beliefs) deriving subsequent intended inferences that, if confirmed by mutually expected goals, communicate speaker intent.

One might still argue that the reasoning involved in processing these indirect speech acts has become short-circuited into a convention of language usage. According to such an account, "conscious" inference is unnecessary for utterances with conventional forms, such as "Can you do X?" For such utterances, the argument goes, people simply "know" such utterances are conventional requests. In contrast, the plan-based theory would appear to propose a Baroque way of uncovering the speaker's intention.

Apart from the lack of evidence that, for example, the "there's a" construction has become conventionalized, accounts of indirect speech acts based solely on convention are inadequate. A conventional account cannot handle the many creative uses of indirection (e.g., "It's cold in here"), nor the case of intended literal interpretations of conventionally indirect speech acts (such as asking a companion on a lifeboat, "Can you swim to shore?"). The plan-based theory is suited to such cases. On the other hand, the indirect speech acts that are usually regarded as conventional can be handled within the plan-based theory in a comparably efficient way, because inference paths can be precomputed from surface speech acts characterizing utterance forms (rather than always being derived from first principles, as various critiques assume). "Bottom-up" *derived rules* can be used in concert with the more general-purpose rules, much as lemmas are used in a proof (Cohen and Levesque 1980).[30] Such an "ability is needed to account for examples such as "Can you reach the hammer", that cannot be handled by conventional methods alone (which would only be able to derive a request to reach the hammer, rather than as a request to pass it).

Scripts are often viewed as short-cuts for more general processing. The plan-based position advocated here conforms to the intuition that scripted situations should give rise to simple processing. By combining strong top-down expectations with bottom-up derived rules based on utterance form, utterances can be interpreted with minimal inference. However, because the plan-based approach incorporates both short-cuts and general mechanisms for reasoning about speaker intent, the theory applies equally well in nonscripted situations.

In summary, a script analysis, if appropriately detailed and formalized, supports the position that speaker intent plays a role in noun phrase interpretation, and that referent identification needs to be reasoned about in the same way as other acts. Essentially, the argument states that scripts are frozen plans, that apprentices inferred these plans, and that the plans indicated that the apprentice should perform actions of referent identification.

The next section shows that the data in this experiment cause difficulty for Searle's analysis of referring. Furthermore, it shows that the present analysis can be extended to cover those cases of referring for which Searle's is applicable.

8 Referring as Requesting

Searle (1969) has argued forcefully that referring is a speech act; that people refer, not just expressions. This section considers what kind of speech act referring might be. I propose a generalization of Searle's "propositional" act of referring that treats it as an illocutionary act, a request, and I argue that a special level of propositional acts for referring is unnecessary.

The essence of the argument is as follows: First, I consider Searle's definition of the propositional act of referring (which I term the PAA, for Propositional Act Account). This definition is found to be inadequate to deal with various utterances in discourse used for the sole purpose of referring. Although the relevance of such utterances to the propositional act has been defined away by Searle, it is clear that any comprehensive account of referring should treat them. I show that the act of requesting referent identification, which I term the *illocutionary act analysis* (IAA) satisfies Searle's conditions for referring, yet also captures utterances that the PAA cannot. The converse position is then examined: Can the IAA capture the same uses of referring expressions as the PAA? If one extends the perceptually based notion of referent identification to include Searle's concept of identification, then by associating a complex propositional attitude to one use of the definite determiner, a request can be derived. The IAA thus handles the referring use of definite noun phrases with independently motivated rules. Referring becomes a kind of requesting. Hence, the propositional act of referring is unnecessary.

[30] These derived rules are akin to Morgan's (1978) "short-circuited implicatures".

8.1 Referring as a propositional speech act

Revising Austin's (1962) locutionary/illocutionary dichotomy, Searle distinguishes between illocutionary acts (IAs) and propositional acts (PAs) of referring and predicating. Both kinds of acts are performed in making an utterance, but propositional acts can only be performed in the course of performing some illocutionary act.

Let us consider Searle's rules for referring, the PAA. A speaker, S, "successfully and non-defectively performs the speech act of singular identifying reference" in uttering a referring expression, R, in the presence of hearer, H, in a context, C, if and only if:

1. Normal input and output conditions obtain.
2. The utterance of R occurs as part of the utterance of some sentence (or similar stretch of discourse) T.
3. The utterance of T is the (purported) performance of an illocutionary act.
4. There exists some object X such that either R contains an identifying description of X or S is able to supplement R with an identifying description of X.
5. S intends that the utterance of R will pick out or identify X to H.
6. S intends that the utterance of R will identify X to H by means of H's recognition of S's intention to identify X, and he intends this recognition to be achieved by means of H's knowledge of the rules governing R and his awareness of C.
7. The semantical rules governing R are such that it is correctly uttered in T in C if and only if conditions 1-6 obtain." (Searle 1969, pp. 94-95.)

Conditions 2 and 3 are justified as follows:

> Propositional acts cannot occur alone; that is one cannot just [emphasis in original – PRC] refer and predicate without making an assertion or asking a question or performing some other illocutionary act.... One only refers as part of the performance of an illocutionary act, and the grammatical clothing of an illocutionary act is the complete sentence. An utterance of a referring expression only counts as referring if one says something. (*Ibid,* p. 25.)

The essence of Conditions 4 and 5 is that the speaker needs to utter an "identifying description", For Searle, "identification" means ".... there should no longer be any doubt what exactly is being talked about". (*Ibid,* p. 85.) Furthermore, not only should the description be an identifying one (one that would pick out an object), but the speaker should intend it to do so uniquely (Condition 5). Moreover, the speaker's intention is supposed to be recognized by the hearer (Condition 6). This last Gricean condition is needed to distinguish having the hearer pick out an object by referring to it from, for example, hitting him in the back with it.

8.2 Problems for the propositional act account

I have shown that in giving instructions over a telephone, speakers, but not users of keyboards, often make separate utterances for reference and for predication. Frequently, these "referential utterances" take the form of existential sentences, such as "Now, there's a black O-ring", Occasionally, speakers use questioned noun phrases – "OK, now, the smallest of the red pieces?" The data present two problems for the PAA.

8.2.1 Referring as a sentential phenomenon

Conditions 2 and 3 require the referring expression to be embedded in a sentence or "similar stretch of discourse" that predicates something of the referent as part of the performance of some illocutionary act. However, it is obvious that speakers can refer by issuing isolated noun phrases or prepositional phrases. Because speakers performed illocutionary acts in making these utterances, then, according to Conditions 2 and 3, there should be an act of predication, either in the sentence or the "similar stretch of discourse". For example, consider the following dialogue fragment:

1. "Now, the small blue cap we talked about before?"
2. "Uh-huh"
3. "Put that over the hole on the side of that tube...."

The illocutionary act performed by uttering phrase (1) is finished and responded to in phrase (2) before the illocutionary act performed in phrase (3) containing the predication "put" is performed. The appeal to a sentence or stretch of discourse in which to find the illocutionary act containing the propositional act in (1) is therefore unconvincing. The cause of this inadequacy is that, according to Searle, to perform an illocutionary act, an act of predicating is required, and the predicate must be uttered (Searle, *op. cit.,* pp. 126-127). Hence, there is no appeal to context to supply obvious predications. Likewise, there is no room for context to supply an obvious focus of attention. Unfortunately, we can easily imagine cases in which an object is mutually, but nonlinguistically, focused on (e.g., when Holmes, having come upon a body on the ground, listens for a heartbeat, and says to Watson: "Dead"). In such a case, we need only predicate. Thus, the requirement that the act of reference be jointly located with some predication in a sentence or illocutionary act is too restrictive – the *goals* involved with reference and predication can be satisfied separately and contextually. The point of this paper is to bring such goals to the fore.

8.2.2 Referring without a propositional act

The second problem is that many of the separate utterances issued to secure reference were declarative sentences whose logical form was $\exists\ x\ P(x)$. Consider, for example, "There is a little yellow piece of rubber", and, "it's got a plug in it". However, Searle claims that these utterances contain *no* act of referring (to x). (Searle, *op. cit.*, p. 29.) How then can speakers use them to refer?

The answer involves our analysis of indirect speech acts. Although such declarative utterances can be issued just to be informative, they are also issued as requests that the hearer identify the referent. The analysis of these utterances as requests depends on our positing an action of referent identification.

8.3 Accounting for Searle's conditions on referring

Assume Searle's Condition 1, the "normal I/O conditions." For the reasons outlined above, do not assume Conditions 2 and 3. Now, clearly, a speaker's planning of a request that the hearer identify the referent of some description should comply with the rules for requesting; the speaker is trying to achieve one of the effects of the requested action (i.e., IDENTIFIED-REFERENT) by way of communicating (in the Gricean sense) his intent that the hearer perform the action, provided that it is shared knowledge that the hearer can do the action. The last condition is true if it is shared knowledge that the precondition to the action holds, which includes Searle's existential Condition 4. Searle's Condition 5 states that the speaker intends to identify the referent to the hearer. This condition is captured in the IAA by the hearer's recognizing that the speaker intends to achieve the effect of the referent identification act, IDENTIFIED-REFERENT. Finally, Searle's Gricean intent recognition, Condition 6, takes hold in the same way that it does for other illocutionary acts, namely by virtue of a "feature" of the utterance (e.g., utterance mood) that is correlated with a complex propositional attitude. This attitude becomes the basis for subsequent reasoning about the speaker's plans. In summary, Searle's conditions can be accounted for by simply positing an action that the speaker requests and that the hearer reasons about and performs.

8.4 Extending the analysis

So far, the IAA and PAA are complementary. They each account for different aspects of referring. The IAA characterizes utterances whose sole purpose is to secure referent identification, and the PAA characterizes the use of referring phrases within an illocutionary act. I now proceed to show how the IAA can subsume the PAA.

Searle argues that one use of the definite article in uttering an NP is to *indicate* the speaker's intention to refer uniquely. Moreover, from Condition 5, this intention is supposed to be recognized by the hearer. We can get this effect by correlating the expression in Figure 5 with the definite determiner, where **(DONE Agt Act P)** is true if the **Agt** has done **Act**, thereby producing the state of affairs **P**. Think of this entire expression as being a pragmatic "feature" of a syntactic constituent, as in current linguistic formalisms. When this expression is applied to a descriptor (supplied from the semantics of the NP), we have a complete formula that becomes the seed for deriving a request. Namely, if the hearer believes that the uttering of the determiner was intentional (which, of course, he does), then the hearer believes that the speaker wants him to think there is a unique object that speaker wants him to pick out. If it is mutually believed, the hearer can do it (i.e., the preconditions to the referent identification act hold, and the hearer knows how to do it by decomposing the description into a plan of action), the hearer believes that the speaker's goal is that he believe of some object that the speaker's goal is that he have picked it out. Hence, a request can be derived.[31] Thus, for the perceptual case, the IAA subsumes the PAA.

Assume that instead of just considering the act of identification in its perceptual sense, we adopt Searle's concept – namely that ".... there should no longer be any doubt what exactly is being talked about." Identification in this sense is primarily a process of establishing a co-referential link between the description in question and some other whose referent is in some way known to the hearer. However, we again regard identification as an act that the hearer performs, not something the speaker does to/for a hearer. If an analysis of this extended notion can be made similar in form to the analysis of the perceptual identification act, then the IAA completely subsumes the PAA. Because both accounts are equally vague on what constitutes identification (as are, for that matter, all other accounts of which I am aware), the choice between them must rest on other grounds. The grounds favoring the identification request analysis include the use of separate utterances and illocutionary acts for its analysis of referring, and the independently motivated satisfaction of Searle's conditions on referring.

8.5 Searle vs. Russell

Using the propositional act of referring, Searle argues against Russell's (1905) theory of descriptions, which holds that the uttering of an expression "the ϕ" is equiv-

[31] The analysis of referring in Cohen (1984) makes use of a theory of communication of Cohen and Levesque (1980, in preparation) that does not require the recognition of illocutionary acts. Instead, IA's are derived as theorems about the speaker's goals. The analysis of requesting in this theory would state that the request theorem is applicable when Searle claims that an act of referring is performed, and, as we have seen, at other times as well. Thus, the hearer does not have to recognize each referring act as a request; each referring act merely has to be characterizable as one.

```
λ D (BEL Hearer
        (WANT Speaker
                (BEL Hearer
                            ∃ ! X (WANT Speaker
                                        (DONE Hearer
                                                IDENTIFY-REFERENT
                                                        (Hearer, D),
                                                IDENTIFIED-REFERENT
                                                        (Hearer, D, X))))))]
```

Figure 5. *Pragmatic feature correlated with a definite determiner.*

alent to the assertion of a uniquely existential proposition, "There is a unique ϕ". Thus, when reference fails, it is because the uniquely existential proposition is not true. Searle claims instead that the existence of the referent is a precondition to the action of referring. In referring to X, we do not assert that X exists any more than we do in hitting X (Searle, *op. cit.,* p. 160.) However, the precondition is necessary for successful performance. Searle's argument against this theory essentially comes down to:

> .... It [Russell's theory] presents the propositional act of definite reference, when performed with definite descriptions ... as equivalent to the illocutionary act of asserting a uniquely existential proposition, and there is no coherent way to integrate such a theory into a theory of illocutionary acts. Under no condition is a propositional act identical with the illocutionary act of assertion, for a propositional act can only occur as part of some illocutionary act, never simply by itself.
>
> (Searle, *op. cit.,* p. 15.)

There are two difficulties with this argument. First, the requirement that acts of referring be part of an illocutionary act was shown to be unnecessarily restrictive. Second, there is a way to assimilate the assertion of an existential proposition – an act that Searle claims does not contain a referring act – into an analysis of illocutionary acts, namely as an indirect request for referent identification. However, because an assertion of a uniquely existential proposition may fail to convey an indirect request for referent identification (just as uttering, "It's cold in here", may fail to convey an indirect request), Searle's argument, though weakened, still stands.

8.6 Summary

There are a number of advantages for treating referent identification as an action that speakers request, and thus for treating the speech act of referring as a request. The analysis not only accounts for data that Searle's analysis cannot, but also predicts each of Searle's conditions for performing the act of singular identifying reference, while allowing for appropriate extension into a planning process. If we extend the perceptual use of referent identification to Searle's more general concept of identification, and we correlate a certain (Gricean) propositional attitude with the use of definite determiners in a noun phrase, then Searle's analysis is subsumed by the act of requesting referent identification. The propositional act of referring is therefore unnecessary.

9 Conclusions

This paper has had five objectives: to develop a methodology for analyzing discourse pragmatics; to apply it in comparing spoken and keyboard discourse; to explore the differences in utterance function across modes (particularly in the pragmatics of reference); to evaluate the adequacy of a plan-based theory of communication for analyzing discourse; and to compare the resulting analysis with Searle's. I shall first summarize the empirical findings and the theory's adequacy, and then discuss implications for computational linguistics.

9.1 Summary of findings

Spoken and keyboard-based instructional discourse, even used for the same ends, differ in structure and in form. Telephone conversation about object assembly is dominated by explicit requests to identify objects satisfying descriptions. This paper makes no attempt to explain why it is that speakers break up the referring and predicating functions, but users of keyboards do not. Many intuitive explanations come to mind – channel bandwidth, memory limitations on the speaker or on the hearer, etc. Rather than attempt more detailed investigation of the causes of the phenomena, I concentrate instead on applying a plan-based theory to derive the coders' analyses of the speakers' intentions.

Only rarely are identification requests performed directly. The plan-based theory was shown to capture many of the indirect requests (69%). However, it needs to be extended in many ways. In particular, the theory does not yet account for the inferring of expected goals from analyses of the communication task, the physical setting, and the modality constraints. It cannot relate perceptual actions to the act of searching for something, and it cannot capture the parallelism evident in speakers' prosodically questioning referent identification in the course of issuing an imperative utterance.

I have argued that the need to process indirect identification requests requires hearers to reason about the speaker's intention that the hearer perceptually identify the referents of various descriptions. This reasoning process involves determining how (the hearer's performing) an *action* of referent identification might fit into the speaker's plans. The treatment of referent identification as an action that speakers request not only accounts for the data, but also predicts each of Searle's conditions for performing the act of singular identifying reference while it allowing for appropriate extension into a planning process.

The promissory note introduced by this approach is to show how the same kind of plan-based reasoning used in analyzing indirect speech acts can take hold when a hearer realizes he cannot, and was not intended to, identify the referent of a description. That is, plan-based reasoning should explain how a hearer might decide that the speaker's intention cannot be what it appears to be (based on the intent correlated with the use of a definite determiner), leading him, for example, to decide to treat a description attributively (Donnellan 1960). Moreover, such reasoning should be useful in determining intended referents, as Ortony (1978) has argued.

To cash in this promissory note, one needs to be specific about speaker intentions for other uses of noun phrases. This will be no easy task. One difficulty will be to capture the distinction between achieving effects on a hearer, and doing so communicatively (i.e., in the Gricean way). Thus, for example, a hearer cannot comply with the illocutionary act, "Quick, don't think of an elephant", because there seems to be an "automatic" process of "concept activation" (Appelt 1981). Achieving effects noncommunicatively, without the recognition of intent, may be central to some kinds of reference. In such cases, speakers would be able to identify referents *for* a hearer. If this held for singular identifying reference, then there could be grounds for a propositional act. However, we might have to give up the Gricean Condition 5, which I suspect Searle would not want to do.

Although it has been demonstrated that the action of perceptual identification differs from other treatments of reference in computational linguistics, perceptual identification has not yet been formalized in a precise way. A formalism should indicate when identification is necessary, how it relates to the performance of physical action, and how the perceptual actions that comprise it are related to the structure of the descriptions. Nonetheless, I hope to have demonstrated the importance of formalizing this action.

The methodology of attempting to explain the discourse analyses of two observers is clearly good theoretical hygiene, but is difficult to implement in practice. Care must be taken to assure that coders are not asked to make too many judgments at once (they get confused), or to make subtle judgments. For example, coders were not able to differentiate questions from requests for confirmation reliably, even though they were provided with formal definitions for the communicative actions and the intuitive distinctions were clear. In fact, my providing the coders with formal definitions may have been counterproductive because the definitions did not necessarily mirror people's commonsense distinctions.

One lesson to be learned from the difficult experience of getting others to "code" the illocutionary acts in a dialogue is that because it is so difficult, perhaps it is not done by the participants. It is certainly possible that conversants engage in dialogue without being able to specify *precisely,* using illocutionary verbs, just which illocutionary acts were performed. Because elsewhere (Cohen and Levesque 1980, in preparation) we have developed formalisms for communication that do not require the identification of illocutionary acts, it may be unnecessary to require the hearer to do so. Therefore, empirical support is needed for the often presupposed position that hearers must identify illocutionary acts.

9.2 Implications for computational linguistics

The simple communication task analyzed here involved primarily the performing of requests. As such, it should fall within the scope of computational linguistic techniques. However, no natural language system that I know can handle the discourse structure of the Telephone data. Of course, one might be able to develop a system that could handle just the data found here. But, we would value more a system (or theory, for that matter) that handles such data because of general principles it embodies. It is for this reason that I have applied a general purpose plan-based theory.

An extrapolation of the empirical results results suggests that if robots are to be instructed with spoken natural language, they are likely to encounter indirect identification requests. If their language processing is based on inferring and responding to the speaker's intent, then intent recognition will have to be applied to the act of identifying a referent. However, given sufficient restrictions on the domain of discourse and the robot's capabilities, an implementation of the plan-recognition process might simplify the general case. I have suggested that derived inference rules coupled with expectations can serve adequately. A formal foundation for such derived rules for intent interpretation can be found in Cohen and Levesque (1980, in preparation), and a prototype implementation that uses both general and derived rules for interpreting speaker intent is described in Brachman et al. (1979).

Additional problem areas suggested by this research include developing formal and computational models of the generation of useful descriptions, and getting machines to plan reference and predication separately.[32] However, if given more resources, systems should be able to "optimize" referring and predicating plans into single utterances. Conversely, systems ought to be able to reason about the speakers' uses of descriptions – for identification, correcting previous misidentifications, attribution, etc.

The empirical findings of this study must be interpreted with three cautionary notes. First, the category of identification requests is specific to discourse situations in which the topics of conversation include objects physically present to the hearer. If the conversation is not about manipulating concrete objects, different pragmatic inferences could be made, even though the same surface forms might be used. Second, not all natural language communication between person and machine are accurately captured by the Telephone and Keyboard conditions. For example, conversations about the contents of the system's display scope (Brachman et al. 1979, Winograd 1972) might share some aspects of the Face-to-Face condition (especially the experts' use of sentence fragments to correct the apprentices' mistakes). Thus, the generality of these findings will only be established when conversations in other discourse situations are analyzed. Third, it should be realized that 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. The question of how people will speak to machines probably cannot be settled until good speech-understanding systems have been developed. Nevertheless, in building future speech-understanding systems, it may be unwise to underestimate the frequency of indirect speech acts in spoken discourse.

Finally, I observe again that when computational linguistic techniques have been developed based on a corpus of dialogues, most often those dialogues have been conducted through keyboard interaction. However, it is clear from the results of this study that keyboard communication is distinctly different from other modalities. In addition to differences from telephone communication, a cursory examination of the handwritten transcripts reveals that keyboard communication is markedly different in structure from written communication. Whereas experts in keyboard mode rarely use identification requests, writers use them frequently, in both direct and indirect forms. Furthermore, writers often performed all identification requests first, and labeled each of the objects for future reference (much as authors of published assembly instructions do). Keyboard interaction, in its emphasis on optimal packing of information into the smallest linguistic "space", appears to be a mode of communication that alters the normal organization of discourse. We should thus be wary of our theories' and techniques' coverage if they are to extended to other modalities of communication.

10 Acknowledgements

This work would not have been possible without the help of Scott Fertig and Kathleen Starr, who spent countless hours coding the transcripts. Many thanks also to Rob Tierney for collaboration in organizing and conducting the experiments. I would like to thank Zoltan Uzhelyi for videotaping, Debbie Winograd, Larry Shirey, and Julie Burke for helping to conduct the experiments, and Marion Hazen, Elsie Chappell, Joan Hirschkorn, Cindy Hunt, Mike Nivens, Ken Olum, and Norma Peterson for text and transcript preparation. Thanks also go to Marilyn Adams, Doug Appelt, Ron Brachman, Chip Bruce, Julie Burke, Martin Kay, Hector Levesque, Sharon Oviatt, Ray Perrault, Andee Rubin, Candy Sidner, Ed Smith, and Rob Tierney for valuable discussions.

References

Allen, J.F. 1979 (January) A Plan-based Approach to Speech Act Recognition. Technical Report 131, Department of Computer Science, University of Toronto.

Allen, J.F. and Perrault, C.R. 1980 Analyzing Intention in Dialogues. *Artificial Intelligence* 15(3): 143-178.

Appelt, D. 1981 (December) Planning Natural Language Utterances to Satisfy Multiple Goals. Ph.D. Thesis, Stanford University, Stanford, California.

Asher, S.R. 1979 Referential Communication. In Whitehurst, G.J. and Zimmerman, B.Z., Eds., *The Functions of Language and Cognition.* Academic Press, New York, New York.

Austin, J.L. 1962 *How to Do Things with Words.* Oxford University Press, London.

Brachman, R.; Bobrow, R.; Cohen, P.; Klovstad, J.; Webber, B.L.; and Woods, W.A. 1979 (August) Research in Natural Language Understanding. Technical Report 4274, Bolt Beranek and Newman Inc.

Bruce, B.C. 1981 Natural Communication Between Person and Computer. In Lehnert, W. and Ringle, M., Eds., *Strategies for Natural Language Processing,* Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Bruce, B.C. 1983 Belief Systems and Language Understanding. In *Trends in Linguistics, Studies and Monographs 19: Computers in Language Research 2.* Walter de Gruyter and Co., New York, New York.

Bruce, B.C. and Newman, D. 1978 Interacting Plans. *Cognitive Science* 2(3): 195-233.

Bruce, B., and Schmidt, C.F. Episode Understanding and Belief Guided Parsing. Presented at the Association for Computational Linguistics Meeting at Amherst, Massachusetts.

Burke, J.A. 1982 An Analysis of Intelligibility in a Practical Activity: The Role and Relationship of Discourse and Context. Ph.D. Thesis, Dept. of Speech Communication, University of Illinois.

Chafe, W.L. 1982 Integration and Involvement in Speaking, Writing, and Oral Literature. In Tannen, D., Ed., *Spoken and Written Language: Exploring Orality and Literacy.* Ablex Publishing Corporation, Norwood, New Jersey.

Chapanis, A.; Ochsman, R.B.; Parrish, R.N.; and Weeks, G.D. 1972 Studies in Interactive Communication: I. The Effects of Four Communication Modes on the Behavior of Teams during Cooperative Problem Solving. *Human Factors* 14: 487-509.

[32] See Appelt (1981) for a system that operates along these lines.

Chapanis, A.; Parrish, R.N.; Ochsman, R.B.; and Weeks, G.D. 1977 Studies in Interactive Communication: II. The Effects of Four Communication Modes on the Linguistic Performance of Teams during Cooperative Problem Solving. *Human Factors* 19(2): 101-125.

Clark, H.H. and Wilkes-Gibbs, D. Referring as a Collaborative Process. Unpublished ms.

Cohen, P.R. On Knowing What to Say: Planning Speech Acts. Ph.D. Thesis and Technical Report No. 118, Department of Computer Science, University of Toronto, Toronto.

Cohen, P.R. 1984 Referring as Requesting. *Proceedings of COLING84,* Stanford, California, 207-211.

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

Cohen, P.R. and Levesque, H.J. (in preparation) Speech Acts as Summaries of Shared Plans.

Cohen, P.R. and Perrault, C.R. 1979 Elements of a Plan-based Theory of Speech Acts. *Cognitive Science* 3(3): 177-212.

Dickson, W.P. 1981 *Childrens's Oral Communication Skills.* Academic Press, New York, New York.

Donnellan, K. 1960 Reference and definite description. *The Philosophical Review* 75: 281-304.

Dore, J.; Gearhart, M.; and Newman, D. 1978 The Structure of Nursery School Conversation. In Nelson, K., Ed., *Children's Language. Vol I.* Gardner Press, New York, New York, 337-396.

Evans, D. 1981 (December) Situations and Speech Acts: Toward a Formal Semantics of Discourse. Ph.D. Thesis, Department of Linguistics, Stanford University.

Fertig, S. Miscommunication in Discourse. Unpublished B.A. Thesis, Hampshire College, Amherst, Massachuseets.

Genesereth, M.R. 1978 (September) Automated Consultation for Complex Computer Systems. Ph.D. Thesis, Department of Computer Science, Division of Applied Sciences, Harvard University.

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

Grosz, B.J. 1977 (July) The Representation and Use of Focus in Dialogue Understanding. Technical Report 151, Artificial Intelligence Center, SRI International.

Hayes, P.J. and Mouradian, G.V. 1981 Flexible Parsing. *American Journal of Computational Linguistics* 7(4): 232-242.

Hindle, D. 1983 Deterministic Parsing of Syntactic Non-fluencies. *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics,* Cambridge, Massachusetts, 123-128.

Hintikka, J. 1969 Semantics for Propositional Attitudes. In Davis, J.W. et al., Eds., *Philosophical Logic.* D. Reidel Publishing Co., Dordrecht, Holland.

Hoeppner, W.; Morik, K.; and Marburger, H. 1984 (May) Talking It Over: The Natural Language Dialog System HAM-ANS. Technical report ANS-26, Research Unit for Information Science and Artificial Intelligence, University of Hamburg.

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

Krauss, R.M. and Weinheimer, S. 1966 Concurrent Feedback, Confirmation, and the Encoding of Referents in Verbal Communication. *Journal of Personality and Social Psychology* 4: 343--346.

Kroch, A.S. and Hindle, D. 1982 On the Linguistic Character of Non-standard Input. *Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics,* Toronto, Canada, 161-163.

Kwasny, S.C. and Sondheimer, N.K. 1981 Relaxation Techniques for Parsing Ill-formed Input. *American Journal of Computational Linguistics* 7(2): 99-108.

Labov, W. and Fanshel, D. 1977 *Therapeutic Discourse.* Academic Press, New York, 1977.

Levin, J.A. and Moore, J.A. 1977 Dialogue Games: Metacommunication Structures for Natural Language Interaction. *Cognitive Science* 1(4): 395-420.

Mann, W.C.; Carlisle, J.H.; Moore, J.A.; and Levin, J.A. 1977 (January) An Assessment of Reliability of Dialogue-annotation Instructions. Technical Report ISI/RR-77-54, Information Sciences Institute.

Mann, W.; Moore, J.; and Levin, J. 1977 A Comprehension Model for Human Dialogue. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence,* Cambridge, Massachusetts.

Moore, R.C. 1980 (October) Reasoning about Knowledge and Action. Technical Note 191, Artificial Intelligence Center, SRI International.

Morgan, J.L. 1978 Two Types of Convention in Indirect Speech Acts. In Cole, P., Ed., *Syntax and Semantics, Volume 9: Pragmatics,* Academic Press, New York, New York, 261-280.

Ochs, E. 1979 Planned and unplanned discourse. In Givon, T., Ed., *Syntax and Semantics, Volume 12: Discourse and Syntax,* Academic Press, New York, New York, 51-80.

Ochs, E.; Schieffelin, B.B.; and Pratt, M.L. 1979 Propositions Across Utterances and Speakers. In Ochs, E., and Schieffelin, B. B., Eds., *Developmental Pragmatics,* Academic Press, New York, New York.

Ortony, A. 1978 Some Psycholinguistic Constraints on the Construction and Interpretation of Definite Descriptions. *Proceedings of the Second Conference on Theoretical Issues in Natural Language Processing,* Urbana, Illinois, 73-78.

Perrault, C.R. and Allen, J.F. 1980 A Plan-based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics* 6(3): 167-182.

Perrault, C.R. and Cohen, P.R. 1981 It's for Your Own Good: A Note on Inaccurate Reference. In Joshi, A.; Sag, I.; and Webber, B., Eds., *Elements of Discourse Understanding.* Cambridge University Press, Cambridge, Massachusetts.

Reichman, R. 1981 Plain-speaking: A Theory and Grammar of Spontaneous Discourse. Ph.D. Thesis, Department of Computer Science, Harvard University, Cambridge, Massachusetts.

Robinson, A.E.; Appelt, D.E.; Grosz, B.J.; Hendrix, G.G.; and Robinson, J.J. 1980 (March) Interpreting Natural-language Utterances in Dialogs about Tasks. Technical Note 210, Artificial Intelligence Center, SRI International.

Rubin, A.D. 1980 A Theoretical Taxonomy of the Differences Between Oral and Written Language. In Spiro, R.; Bruce, B.; and Brewer, W., Eds., *Theoretical Issues in Reading Comprehension,* Lawrence Erlbaum Assocs., Hillsdale, New Jersey.

Russell, B. 1905 On denoting. *Mind* 14: 479-492.

Sacerdoti, E.D. 1975 (August) A Structure for Plans and Behavior. Technical Note 109, Artificial Intelligence Center, SRI International.

Schank, R. and Abelson, R. 1977 *Scripts, Plans, Goals, and Understanding.* Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Schmidt, C.F. 1975 Understanding Human Action. *Proceedings of Conference on Theoretical Issues in Natural Language Processing,* Cambridge, Massachusetts.

Schmidt, D.F.; Sridharan, N.S.; and Goodson, J.L. 1979 The Plan Recognition Problem: An Intersection of Artificial Intelligence and Psychology. *Artificial Intelligence* 10: 45-83.

Searle, J.R. 1969 *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge.

Shatz, M. and Gelman, R. 1973 The Development of Communication Skills: Modifications in the Speech of Young Children as a Function of Listener. Monographs of the Society for Research in Child Development.

Sidner, C.L. 1979 (June) Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. Technical Report 537, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Sidner, C.L. 1983 The Pragmatics of Non-anaphoric Noun Phrases. In Research in Knowledge Representation for Natural Language Understanding: Annual Report, 9/1/82–8/31/83, Bolt Beranek and Newman Inc., Cambridge, Massachusetts.

Sidner, C.L., Bates, M.; Bobrow, R.J.; Brachman, R.J.; Cohen, P.R.; Israel, D.J.; Webber, B.L.; and Woods, W.A. 1981 (November)

Research in Knowledge Representation for Natural Language Understanding. Annual Report 4785, Bolt, Beranek and Newman Inc.

Sidner, C. and Israel, D. 1981 Recognizing Intended Meaning and Speaker 's Plans. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence,* Vancouver, B. C.

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

Stoll, F.C.; Hoecker, D.G.; Krueger, G.P.; and Chapanis, A. 1976 The Effects of Four Communication Modes on the Structure of Language Used During Cooperative Problem Solving. *The Journal of Psychology* 94(1): 13-26.

Thompson, B. 1980 Linguistic Analysis of Natural Language Communication with Computers. *Proceedings of COLING-80,* Tokyo, 190-201.

Tierney, R.J.; LaZansky, J.; Raphael, T.; and Cohen, P.R. 1983 Author's Intentions and Readers' Interpretations. In Tierney, R.J.; Anders, P.; and Mitchell, J.N.; Eds.; *Understanding Readers' Understandings.* Lawrence Erlbaum Assoc., Hillsdale, N. J., 1983.

Walker, D., Ed. 1978 *Understanding Spoken Language.* Elsevier North-Holland, New York New York.

Webber, B.L. 1978 (May) A Formal Approach to Discourse Anaphora. BBN Report 3761, Bolt Beranek and Newman Inc.

Weischedel, R.M. and Black, J.E. 1980 Responding Intelligently to Unparsable Inputs. *American Journal of Computational Linguistics* 6(3): 97-109.

Wilensky, R. 1978 Understanding Goal-based Stories. Research Report 140, Department of Computer Science, Yale University, New Haven, Connecticut.

Wilkes-Gibbs, D. How to Do Things with Reference: The Function of Goals in Determining Referential Choice. Unpublished ms.

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

Woods, W.; Bates, M.; Brown, G.; Bruce, B.; Cook, C.; Klovstad, J.; Makhoul, J.; Nash-Webber, B.; Schwartz, R.; Wolf, J.; and Zue, V. 1976 Speech Understanding Systems – Final Technical Progress Report. Technical Report 3438, Bolt Beranek and Newman Inc.

Appendix A

Instructions for the Expert

We are studying communication between individuals. We want to examine how an individual effectively communicates a set of instructions to another. The purpose of this experiment is to observe and document the variety and similarities among different communicative styles.

If you decide to participate, you will be randomly assigned to another individual taking part in the experiment, and to communication modality. Each pair is to cooperate in building a water cannon pump. You will have been previously trained to build the pump, but will not be allowed to touch any of the parts during the experiment. Your partner will do the actual building.

You should be aware that your partner will know very little about the task. You must be sure to explain what the task is, make sure they build the pump, and check to see that the pump functions correctly.

The schedule for your role in the experiment is as follows. Today you will read the instructions for building the water pump and then practice building it. After about 20 minutes, we will ask you to instruct one of us. If there are no problems with this practice run, then you are free to leave and tomorrow you will come in and instruct your partner. If you do have difficulty or have any doubts about your ability to remember till tomorrow how to assemble the pump, then we want you to stay and practice for at least another ten minutes.

The communication will take place in one of the following modes: face-to-face, telephone, teletype, audiotape, or written. You will learn the specifics of your mode in the next session. Depending on the mode of communication, you may be recorded on video or audio tape. These tapes will be used for the collection of data and not for any other purpose.

All fellow subjects will be adult university students.

If you agree to these conditions, please sign below.

Instructions for Building a Water Cannon

Building a Water.Cannon:

1. Plug the hole in the bottom of the plunger with the plunger plug.
2. Insert the plunger into the main tube. The red handle of the plunger should extend from the non-threaded end of the main tube.
3. Press the blue tube cap down onto the main tube so it fits firmly.
4. Drop the O-ring into the tube base.
5. Fit the pink base valve onto the top of the tube base. The valve should cover the hole in the base.
6. Fit the feed tube onto the bottom of the base.
7. Screw the tube base onto the main tube.
8. Put the tube cap over the upper outlet of the main tube.
9. Fit the slide valve loosely into the lower outlet of the main tube.

Using the Water Cannon:

1. Place the pump into a tray of water. The pump should be supported by the feed tube.
2. Move the plunger up and down by alternately pushing and pulling on the red handle.
3. Water will be forced out the lower outlet of the main tube, through the spout, through the air chamber, and out through the nozzle.
4. Water will continue to be forced out the nozzle as long as you keep moving the plunger up and down.
5. If nothing happens, check to see that all parts fit tightly and that the valves are properly sealed.

Telephone

Talk to your partner as you would during a normal phone conversation. Your partner will have his/her hands free during the entire conversation and will be able to construct the pump without interruptions.

Your partner will have all the necessary pieces and a tray of water. Again, the task is to assemble the pump and ensure that it works. Your partner does not know anything about the task.

Teletype

Just type as you would on a typewriter. The print will not appear as soon as it is typed in; there will be a small delay. If you experience any long delays in seeing the characters you typed it is probably due to heavy use of the computer. Please bear with it.

Finally, it is possible the computer will stop working during the experiment. Everything before a stopage will be saved and we will attempt to continue the experiment as soon as possible.

Your partner will have all the necessary pieces and a tray of water. Again, the task is to assemble the pump and ensure that it works. Your partner does not know anything about the task.

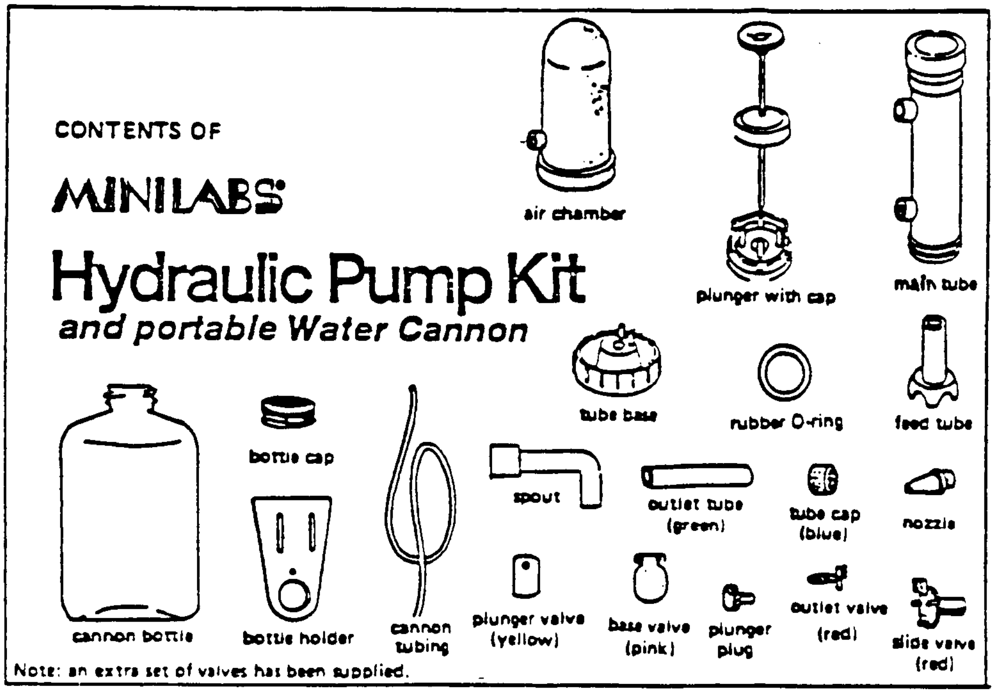
Instructions for the Novice

We are studying communication between individuals. We want to examine how an individual effectively communicates a set of instructions to another. The purpose of this experiment is to observe and document the variety and similarities among different communicative styles.

The communication will take place in one of the following modes: face-to-face, telephone, teletype, audiotape, or written. You will learn the specifics of your mode on the next page. Depending on the mode of communication, you may be recorded on video or audio tape. These tapes will be used for the collection of data and not for any other purpose.

If you agree to these conditions, please sign below.

Exploded Parts Diagram of the Water Pump



Appendix B: Coding Categories

The following is the list of parts and their respective codes. Subparts are indented on a new line after the main part. To avoid confusion, subpart names are used where needed.

Part	Code
Main Tube	MT
Outlet1 (Main Tube)	O1
Outlet2	O2
Plunger [green end]	PL
Plug	PLUG
Rod	ROD
Handle	HANDLE
Top-cap	T-CAP
Outlet1 cap	O-CAP
Tube Base	TB
Valve2 [pink valve]	V2
O-ring	O-RING
Valve3 [red slide valve]	V3
Spout [elbow joint]	SPOUT
Air-chamber	AIR-CH
Nozzle	NOZ
Stand	STAND
Pump [as built so far]	PUMP
Tray	TRAY
Table	TAB
Expert	EXP
Apprentice	APP

Subassemblies

Occasionally, subjects mention a grouping of parts to be regarded as subassemblies, which are then connected together to form the pump. The following are typical ones – what a particular expert groups into one subassembly is up to him/her, so we were not strict on what are the constituents of a subassembly.

- Base-assembly
  BASSM = *{TB,V2,STAND,ORING}*
- Spout-assembly
  SPASSM = *{SPOUT, V3, O2}*
- Main-tube-assembly
  MTASSM = *{MT,PASS,PLUG,T-CAP}*
- Air-chamber-assembly
  ARCHMASSM = *{SPOUT,NOZ,AIR-CH}*

Functions

The following functions can be applied to (the right) parts and yield the appropriate aspects of those parts. So, "Bump" can be applied to the tube base TB to yield the set of 2 bumps protruding from it.

- Bump
- Hole
- Threaded-end
- Nonthreaded-end
- Thin-end
- Fat-end

Actions

The following are the set of actions that are generally used in building the pump.

- PICK-UP(part)
- PUT-DOWN(part)
- PUT-INTO(inserted-part receiving-part)
- PUSH-INTO(inserted-part receiving-part)
- COVER(covering-part covered-part)
- SCREW-TOGETHER(female-part male-part)
- MESH(hole-part bump-part)
- CONNECT(enclosing-part enclosed-part)
- ORIENT(part towards/away-from(part))
- STOP [action]
- UNDO(most-recent-action+2nd-mr-action+...+ last-action-to-undo)
- PUMP [handle]

The next action makes the pump or subassembly from the set of pieces. It was only coded when the expert gave an overview of a number of steps before instructing how to do the substeps.

- ASSEMBLE (PUMP or sub-assembly)
- ACHIEVE (person relation)

ACHIEVE stands for "make [relation] true." ACHIEVE was coded when [relation] was on our list below, but the action that would achieve A was not mentioned.

Relations

The following relations were coded as the content of the categories INFORM and ACHIEVE. Appropriately filled-in with the right parts, a collection of these relations form the goal state of the assembled pump.

- (ACCESSIBLE part)
- (HOLDING part)
- (SUPPORTED part1 part2)
- (INSIDE inserted-part receiving-part)
- (COVERS covering-part covered-part)
- (SCREWED-TOGETHER female-part male-part)
- (MESHES hole-part bump-part)
- (CONNECTS enclosing-part enclosed-part)
- (ORIENTED part)
- (IDENTIFIED part)
- (READY EXP/APP)
- (TIGHT part1 part2) part1 should be tightly connected to part2
- (LOOSE part1 part2) part2 should be loosely connected to part2
- (SEES person object)

- (DOUBTFUL person)
- (WORKING pump)
- (COMPLETED action) implies the proposition stated as the final state of the action holds.

All relations could be negated. This was expressed as "(NOT-[relation])". E.g., (NOT-COMPLETED PICKUP (MT))

Appendix C: Identification Requests in Telephone Modality

The utterances coded as identification requests are presented in italics.

Class A: Exisential Propositions

1: And then– *there should be a tray of water with you?*

1: Okay. Now, if you'll look around in front of you, *there's a little tiny red piece that is like a uh–looks like a fat thumb tack.*

1: and *there's this little tiny like pink plastic thing.*

2: Yeah.

1: Now. *There's another funny little looking red thing, a little teeny red thing that's some– should be somewhere on the desk, that has um–there's like teeth on one end?*

2: Okay.

1: What's next? All right. See that little– *there's a little L-shaped clear plastic.*

2: Yeah.

1: All right. Now. *There is a skinny um– there's a funny blue –blue tube–* it's a skinny blue tube.

2: Mm-hm.

1: *There is one red cap.*

2: Mm-hm.

1: *And a funny like cylinder?*

2: Yeah.

1: All right *Now there's a blue cap that has two little teeth sticking out of the bottom [of it.]*

2: [Mm-hm.]

J: Huh. Okay, first thing, *there's a long cylinder that has a slightly purplish cast to it.*

T: With the two side hoods, yep.

J: Okay. Uh now *there's a little plastic blue cap.*

T: Yep.

J: Uh, the next thing is *there is a blue– looks like a screw cap.*

T: Mm-hm.

J: *There's two little prongs sticking up.*

T: Yeah.

J: Okay, now *there's a black O-ring.*

T: Got it.

J: Okay. In the green thing at the bottom, *there's a hole.*

T: Right.

Put the little red [plug–]

J: [unintelligible] *There's a little red thing you gotta stuff up in there.*

T: Okay. Got it.

J: Okay. Next step.

T: Mm-hm.

J: *There's a little red thing with um prong-like things hangin' out from it.*

T: With what-like things?

J: Okay. Now *there's a clear 90-degree angle piece of plastic.*

T: Right. That–[which fits–]

J: Okay. Now, the next thing is *there's a uh a blue thing with a plastic dome, looks something like somethin' you'd put on uh a to evacuate. I'm talking about– it's the only big piece left really.*

T: Mm-hm.

J: Okay. Onto that *there is a little red nozzle,* and that fits on the side hole coming out of that dome.

T: Okay.

J: Okay. Take the whole mechanism and stand it up into the uh– *I think there's a photographic tray there full of water.*

T: Mm-hm.

S: 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: Okay.

A: Okay. Now *there's a little blue cap?*

J: Yes.

A: Uh-huh. Now, *there's a–a red plastic piece that has four gizmos on it.*

J: Yes.

1: Pick that up, *and it has two projecting blue prongs on it.*

2: Uh-huh.

1: *And it has two holes in it.*

2: Uh-huh.

1: *It's a funny-loo–hollow, hollow projection on one end and then teeth on the other.*

2: Uh-huh.

1: And if you'll look carefully, *one toward the blue end has two holes in it,*

1: *and the other –toward the red end has no holes in it.*

2: Mm-hm.

1: All right. *We have another hole.*

2: Mm-hm.

1: All right. Has a–uh, I mean, [unintelligible] *has a funny projection at the other end, it's like notched.*
2: Mm-hm.

1: -and this um cylinder-like thing, if you look at the bottom *has a hole in it–*
2: Mm-hm.

1: Then, if you'll pick up the tube– *it's kind of a purple color?*
2: Yes.

A: Okay. Now take the littlest red plastic piece– *it's a little um–*
J: Looks like a plug?
A: *stopper*

J: and also pick up what looks like a plunger. *It has [a red end and green–]*
T: [With the red end and a blue–]
J: bottom.
T: Right.

S: Okay? *Now you have two blue caps.*

S: *One very small that's just a push-on cap,*

S: *and the other one is a larger one with threads on it.*
J: Okay.

S: Okay? *Now you have two devices that are clear plastic.*
J: Okay.

S: *One of them has two openings on the outside with threads on the end,* and it's about five inches long. Do you see that?

S: Okay, *the other one is a bubbled piece with a blue base on it with one spout.* Do you see it?

S: *It's just round with a little point.*
J: Yeah

S: Okay, now *you've got a bottom hole still to be filled,* correct?
J: Yeah.

S: Okay. *You have one red piece remaining?*
J: Yeah.

S: Okay. Take that red piece. *It's got four little feet on it?*
J: Yeah.

S: *–and on the bottom of that is a hole,* right?
J: Yeah.

S: Okay? Now, *do you have a–a uh little spout that has a 90 degree turn in it?*
J: Yeah.

A: Okay. Uh let's see. *Got the plunger?* That thing with the metal part and uh–
J: Right.
A: –and the red, blue, and green.
J: Got it.

Okay. Now, insert that entire thing– *the red part of the plunger is your handle.*

A: Now take the plunger and the– the main tube– that's the biggest plastic tube, *and it's got a threaded end and an unthreaded end?*
J: Right.

A: Okay, *now you got a little pink seal with two holes in it?*
J: Yes.

A: Okay. *Now you've got that uh cone-shaped–that sort of mouth-shaped red plastic thing?*
J: Right.

A: Okay, *now all you've got left is that little blue plastic thing.*
J: Right.

A: Okay, *on the bottom of the main tube you've got the big blue cap.*
J: Yes.

A: *And it's got a–it's got a uh peg in the bottom of it.*
J: Right.

A: The main tube, yeah. *The main tube has a blue cap on the bottom,* and also
J: Yes

A: *–a blue cap on the top.*
J: Yes, right.

A: Okay, you've got this blue cap on the bottom. *It's got a peg in it.*
J: Right. Small red peg with four little red things comin' off it.

Class B: Perception-Based

S: Okay. First I want you to uh– *do you see small–three small red pieces?*
J: Y-yes.

S: *Do you see a little ring, a little black rubber ring?*
J: Yeah.

S: Now. *Do you see a little pink plastic piece?*
J: Yeah, yeah.

S: And stick it on the en– onto the uh spout coming out the side. *You see that?*
J: Yeah,

1: *You see a hole?*
2: Uh-huh.

1: *Um, you will see in front of you a bunch of um pieces of plastic.*

1: All right. *And then you'll see the blue cap begins to come down over the tube.*

1: If you'll look at the bottom, *you will see a projection in that cap.*
2: Yeah.

1: *and you see the two projections?*
2: Mm-hmm

2: *You'll see three very small red pieces of plastic.*

Class C: Fragments

1: And–if you're still holding the top– uh the blue–the blue *–that looks like the medicine cap with the peak on it?*
2: Mm-hm.

1:*–and this um cylinder-like thing,* if you look at the bottom,

has a hole in it–
2: Mm-hm.

1: All right. *this very – this very blue um like narrow little funnel*

A: Oh, let's see. Four little things coming off it? Now that was supposed to be–that was supposed to be in the–
J: top.
A: –in the side. *The little red–the red thing with four little things coming off it.*
J: Hey, I think we're there.

S: –pick it up, *and in the bottom of the blue cap on the main tube–*
J: Uh-huh.
S: *–is another hole.*

2: And take a small blue cap and plug the top hole.
1: The top hole?
2: *On the side.*
1: Okay.

2: Now. *In the cap that you plug the bottom of it with–*
1: Mm-hm.

J: Okay. *In the green thing at the bottom,* there's a hole.
T: Right. Put the little red [plug–] &line
S: Okay. 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: Okay.

S: Okay. Now. Now, *the smallest of the red pieces?*
J: Okay.

S: Okay. Now *where the little red valve is* I want you to flip that 90 degree spout over that.
J: Okay.

S: Okay. Now. I want you to take the other tube that now has a little red spout sticking on it–
J: Yeah.
S: –pick it up, and *in the bottom of the blue cap on the main tube–*
J: Uh-huh.
S: –is another hole.

A: Okay, the–the big part goes in the bottom, *and the little part,* that's what you use– uh you should fit into the– the main tube, the bottom.
J: (laughs)

A: Yeah, take the red thing off. That was the wrong instruction.
J: And put the . . .
A: *That water chamber with the blue bottom and the globe top?*
J: Yeah.

J: Hm. (laughs) Where does it go?
A: Uh, *that–that–that–that uh opening in the side?*
J: Yeah.

J: On the red thing?
A: Um no, *just–just in the bottom of the– you know the big blue cap?*
J: Yeah.
A: *In the bottom of the main tube.*
J: Big blue cap
A: Yeah

A: Okay, now stick the elbow joint back on.
J: I got it.
A: Okay, now *on the very bottom– the very bottom of the main tube–*
J: I got it, too.

Class D: Nearly Direct Requests

1: the next thing *you're gonna look for is a uh blue piece– it's–it's uh a fairly large blue piece, and it looks like the cap to a medicine bottle.*

1: *and look at the bottom of the tube that you should be holding*

Class E: "Let's" Requests

1: Okay. Uh keep the [?] *Let's start with a piece that has– it's a metal rod.*
2: Mm-hm.
1: *And it has a green thing on one end–*
2: Yes.
1: *–and a blue and a red*

1: Okay, now (laughs) *let's go back to the original parts that we put together.*

1: *Let's go to the little tiny blue cap.*
2: Okay.

1: All right. Now, *let's go back to that funny little red projection with teeth on the other end.*
2: Okay

1: Okay. Now, *let's go back to that little L-shaped piece of plas–clear plastic that's sticking up–*
2: Mm-hm.

Supplemental NPs

S: Okay, the other one is a bubbled piece with a blue base on it with one spout. Do you see it? About two inches long. *Both of these are tubular.*
J: Okay. Not the bent one.

S: Okay, I want you to take the largest tube, *or actually it's the largest piece of anything,* that has two openings on the side –
J: yeah

S: Take the spout– *the little one that looks like the end of an oil can–*
J: Okay.

S: –and put that on the opening in the other large tube. *With the round top*

A: Now take the plunger and the– the main tube– *that's the biggest plastic tube* and it's got a threaded end and an unthreaded end?
J: Right.

A: Okay. Now, take the big blue stopper that's laying around and take the black ring–
J: The big blue stopper.
A: Yeah, *the big blue stopper* (short pause) *and the black ring.*
J: Yes.

A: Okay, the–the big part goes in the bottom, and the little part, that's what you use– uh you should fit into the– the main tube, *the bottom.*
J: (laughs)

S: Okay. Now, take the larger blue cap, *which is the only cap remaining–*
J: Yeah,

2: And then the elbow goes over that. *The big end of the elbow.*

Appendix D: Sample Dialogues

Sample Telephone Dialogue

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
J: Okay,					
we can start now.					
S: Okay, John,					
you have all the pieces in front of you?					
J: I guess so.					
All of 'em.					
S: Okay.					
First I want you to uh–					
do you see small–three small red pieces?				PLUG, V3, NOZ	
J: Y-yes.					PLUG, V3, NOZ
S: Okay.					
Why don't you take those and separate those out.		PICK-UP(PLUG, V3, NOZ)			
Put those three together.		PUT-DOWN(PLUG, V3, NOZ)			
J: Okay.					
S: Okay?					
Now you have two blue caps.			O-CAP, TB		
One very small that's just a push-on cap,			O-CAP		
and the other one is a larger one with threads on it.			TB		
J: Okay.					O-CAP, TB
S: You see those?				O-CAP, TB	
J: Yes.					O-CAP, TB

S: Put those together and separate.		PICK-UP(O-CAP,TB)
		PUT-DOWN(O-CAP, TB)			

J: Okay.					
S: Okay?					
Now you have two devices that are clear plastic.			MT,AIR-CH		
J: Okay.					MT,AIR-CH
S: One of them has two openings on the outside with threads on the end, and it's about five inches long.			MT		
Do you see that?				MT	
J: Yeah.					MT

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
S: Okay,					
the other one is a bubbled piece with a blue base on it with one spout.			AIR-CH		
Do you see it?				AIR-CH	
About two inches long.					
Both of these are tubular.			MT,AIR-CH		
J: Okay.					
Not the bent one.					
S: No, not the bent one.					
J: Okay.					
S: That's a spout,	SPOUT				
okay?					
Okay,					
I want you to take the largest tube,		PICK-UP(MT)			
or actually it's the largest piece of anything,			MT		
that has two openings on the side–					
J: Yeah.					MT
S: –and threads on the bottom.					
J: Yeah.					MT
S: Do you see it?				MT	
J: Yeah.					MT
S: Okay.					
Take that.		PICK-UP(MT)			
Now there's a thing called a plunger.	PASS			PASS	
It has a red handle on it, a green bottom, and it's got a blue lid.					
J: Okay.					PASS
S: Take that, and starting–		PICK-UP(PASS)			
insert the green end into the top of that large piece that you have in your hand–		PUT-INTO(PL MT)			
J: Okay.					
S: –and push the green thing down until it comes to the threaded end.		PUSH-INTO(PL MT)			
J: Mm.					
It's pretty tight.					

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
S: Tight fit.					
J: Okay.					
S: Is that in?					
J: Yeah.					
S: Now,					
put that–					
snap that blue cap over the top.		COVER(T-CAP MT)			
J: (pause) Okay.					
S: Have that?					
J: Yeah.					
S: Okay.					
Now,					
the small blue cap we talked about before?			O-CAP		O-CAP
J: Yeah.					O-CAP
S: Put that over the hole on the side of that tube–		COVER(O-CAP 01)			MT
J: Yeah.					
S: –that is nearest to the top,					
or nearest to the red handle.					
J: Okay.					
S: You got that on the hole?					
J: Yeah.					
S: Okay.					
Now.					
Now,					
the smallest of the red pieces?			PLUG	PLUG	
J: Okay.					PLUG
S: You see that?				PLUG	
J: Yeah.					
S: It's just round with a little point.			PLUG		
J: Yeah.					PLUG
S: Take that and stick that in the end of the green part of the plunger.		MESH(PL PLUG)			
In the bottom of the green part of the plunger.					
J: Okay.					
S: You see it?				Hole(PL)	
J: Yeah.					Hole(PL)
S: You got it in?					
J: Yeah.					

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
S: Okay.					
Now,					
take the larger blue cap,		PICK-UP(TB)			
which is the only cap remaining–			TB		
J: Yeah.					
S: Do you see a little ring,				O-RING	
a little black rubber ring?					
J: Yeah.					O-RING
S: Slip that down into that cap.		PUT-INTO(O-RING TB)			
J: Okay.					
S: Got it?					
J: Yeah.					
S: Now.					
Do you see a little pink plastic piece?				V2	
J: Yeah, yeah.					V2
S: With two holes?				V2	
J: Yeah.					V2
S: Okay.					
You have your blue cap in front of you?					
J: Yeah.					
S:Setting down with the two little prongs sticking up.					
J: Yeah.					
S: Okay,					
take that little pink plastic piece,		PICK-UP(V2)			
and the two holes in the plastic piece–		MESH(V2 TB)			
J: Mm-hm.					
S: –go over the two little notches.					
J: Does it matter whether the shiny side or the dull side of the pink thing's up?					
S: Pardon me?					
J: Well, one side of the pink thing is shiny, one side is–					
S: No, it doesn't matter.					
J: Okay.					
S: And put it so that it's covering the hole in the bottom of that little cap.		ACHIEVE(COVERS(V2 Hole (TB)))			

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
(pause) Kinda fits hard, doesn't it?					
J: Little bit tight, yeah.					
Okay.					
S: Okay,					
now it's covering the hole in the cap.					
J: Yeah.					
S: Okay.					
Now,					
I want you to screw that cap onto that big air tube that had the threads on it–		SCREW-TOGETHER(MT TB)			
J: Okay.					
S: –that you put the plunger in.					
J: Okay.					
S: Now,					
you have the large tube,					
you have the plunger in the tube,					
you have the little red thing in the bottom of the plunger.					
J: Mm-hm.					
S: You have the blue cap on the upper hole.					
J: Mm-hm.					
S: And you have the big blue cap with the ring and the little plastic thing screwed onto the bottom.					
J: Yeah.					
S: Okay.					
Now.					
Um I want you to take the um–					
okay,					
take uh the					
–now you have two red pieces remaining, right?					
J: Yeah.					
S: Take the spout–		PICK-UP(NOZ)			
the little one that looks like the end of an oil can–			NOZ		

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
J: Okay.					NOZ
S:–and put that on the opening in the other large tube.		CONNECT(NOZ AIR-CH)			
With the round top–					
J: Oh,					
the other large tube.					
Okay.					
S: The other large–					
yeah.					
Put the tube with the plunger aside.		PUT-DOWN(MT)			
J: Okay.					
S: And stick it on the en–		CONNECT(NOZ AIR-CH)			
onto the uh spout coming out the side.					
You see that?				03	
J: Yeah,					03
okay.					
S: You got that on,					
okay.					
J: Yeah.					
S: Um now.					
Now we're getting a little more difficult.					
J: (laughs)					
S: Pick out the large air tube that has the plunger in it.		PICK-UP(MT)			
J: Okay.					
S: And set it on its base,		PUT-DOWN(TB)			
which is blue now,					
right?					
J: Yeah.					
S: Base is blue.					
Okay,					
now					
you've got a bottom hole still to be filled,					
correct?					
J: Yeah.					
S: Okay.					
You have one red piece remaining?					
J: Yeah.					
S: Okay.					

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
Take that red piece.		PICK-UP(V3)			
It's got four little feet on it?				V3	
J: Yeah.					V3
S: And put the small end into that hole on the air tube–		CONNECT(SPOUT AIR-CH)			
on the big tube.					
J: On the very bottom.					
S: On the bottom,					
yes.					
J: Okay.					
S: Okay?					
Now,					
do you have a–a uh little spout that has a 90 degree turn in it?				SPOUT	
J: Yeah.					SPOUT
S: Okay.					
Stick that directly over–		CONNECT(SPOUT---)			
Now wait.					
We put the little red piece in the bottom hole,					
correct?					
J: Yeah,					
on the–					
you mean the bottom hole in the side or the					
–there was a bottom hole in the blue cap.					
S: On the side,					
yes.					
J: On the side,					
okay.					
S: Yes,					
okay?					
One's got a blue cap on it,					
and the other one's got the little red thing in it now.					
J: Yeah.					
S: Okay.					

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
Now where the little red valve is		CONNECT(SPOUT MT)			01
I want you to flip that 90 degree spout over that.					
J: Okay.					
Which way should I point the spout?					
S: The–the spout should point upward.					
J: Upward,					
okay.					
S: Which is toward the red handle of the plunger.					
J: Yeah.					
S: Got that?					
J: Yeah.					
S: Okay.					
Now the little–the little red thing fits in there					
okay.					
J: Yeah.					
S: Okay.					
Now.					
I want you to take the other tube that now has a little red spout sticking on it–		PICK-UP(AIR-CH)			
J: Yeah.					AIR-CH
S: –and on the bottom of that is a hole,				Hole (AIR-CH)	
right?					
J: Yeah.					
S: I want you to fit that over the top of that 90 degree turn.		CONNECT(SPOUT AIR-CH)			
J: Okay.					
S: Okay?					
J: Yeah.					
S: Got that?					
J: Mm-hm.					
S: Now,					
what piece do you have remaining?					
Only the one long bluish colored piece?				STAND	
J: Yeah.					STAND

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
S: Okay.					
Take the entire apparatus,		PICK-UP(PUMP)			
which should all be together now–					
J: Yeah.					
S: –pick it up,					
and in the bottom of the blue cap on the main tube–			Hole(TB)		
J: Uh-huh.					Hole(TB)
S: –is another hole.					
I want you to stick that remaining piece–		CONNECT(STAND TB)			
J: In there?					
S: –in that hole–					
J: Okay.					
S: –with the long piece going in.					
J: Yeah.					
Okay.					
S: Okay.					
Now.					
You have the plunger in and all the holes should be filled.					
J: Uhh.					
S: Are there any remaining pieces?					
J: No.					
S: No.					
Okay.					
Now,					
insert that entire thing–		PUT-INTO(PUMP TRAY)			
the red part of the plunger is your handle.	HANDLE				
Insert the base of the–of the apparatus into the water–					
J: Mm-hm.					
S: –and start pumping that handle.		PUMP[HANDLE]			
Draw up some water,					
like a hypodermic needle.					
J: Okay.					
S: Now,					

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
push it down.					
Does it indeed squirt out the red spout?					
J: Yeah.					
S: Okay. I think we're finished. (END OF TAPE)					

Sample Keyboard Dialogue

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
anyone there?					
hello					
hi					
all right?					
ready					
ok					
N: [what]					
B: to [whom am i] speaking					
N: nicolette					
B: shall we begin?					
N: i'm ready when you are					
B: ok, here goes... 1) take the plunger... and [2)]		Pick-up(PASS)			
N: [ok]					PASS
B: insert it into the non-threaded end of the big tube...		Put-into(PL MT)			
N: ready					
B: fit the blue cap over the tube end		Cover(T-CAP MT)			
N: done					
B: put the little black ring into the large blue cap with the hole in it...		Put-into(O-RING TB)			
N: ok					
B: put the pink valve on the two pegs in that blue cap...		Mesh(V2 TB)			
N: ok					
B: now, put the little blue cap over the hole in the large tube near the plunger handle...		Cover(O-CAP 01)			
N: ready					
B: forgot one thing... use the red thing that looks like a nail to plug the plunger so it will work...					
N: [you mean] the green part					
B: [capeesh] you got it, kid...					
N: great					
B: anyway, put the red piece with the strange					

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
projections LOOSELY into the bottom hole on the main tube. Ok?		Mesh(V3 MT) & Achieve (LOOSE V3 MT)			
N: which hole the bottom one on the side?					
B: right. put the 1/4 inch long ’post’ into the loosely fitting hole...		Mesh (V3 MT)			
N: i don’t understand what you mean					
B: the red piece, with the four tiny projections?				V3	
N: ok					V3
B: just place it loosely [into the]		Mesh (V3 MT) & Achieve (LOOSE V3 MT)			
N: [done]					
B: yes?					
N: yes					
B: place it loosely into the hole on the side of the large tube...					
N: done					
B: very good. See the clear elbow tube?				SPOUT	
N: yes					
B: place the large end over that same place.		Connect (SPOUT MT)			
N: ready					
B: take the clear dome and attach it to the end of the elbow joint...		Pick-up (AIR-CH) Connect (AIR-CH SPOUT)			
N: using the blue attachment part?					
B: right. it’s already attached, so I didn’t mention it. Now, put the red nozzle over the hole in the dome.		Connect (NOZZ AIR-CH)			
N: ok					
B: Almost done now. Screw the blue cap that has the pink valve on it onto the bottom of the main cylinder.		Screw-together (TB MTASSM)			
N: ready					
B: stick the translucent blue stand onto that very cap so that the pump will stand up.		Connect (STAND PUMP) Achieve (SUPPORTED TAB PUMP)			

	Label	Request Action	Request Ident	Request Infif Ident	Inform Complete Ident
N: ok, but it won't stand					
B: then hold onto it...		Achieve (HOLDING PUMP)			
N: ok					
B: i think that that's alll. Test it!		Achieve (WORKING PUMP)			
N: here goes					
B: well???					
N: it works beautifully					
B: thank you, no applause, just money...					
sure thing					
anything else?					
nope, that's all					