[Prof. Ben-Avi]

# CONVERSATION WITH SHRDLU by Terry Winograd

Some theoretical points about the human use of natural language have been stressed, and considerably clarified, by Terry Winograd. His "insights" about natural language were developed in the course of his project to program a robot to answer questions, execute commands, and accept information in an interactive English dialogue with a human being. In other words, to create a natural language "front-end".

The essential feature of Winograd's program, in which it differs radically from Joseph Weizenbaum's ELIZA, is the inclusion of a system of knowledge concerning the assumed universe of discourse, a system on which the program can draw in a reasoned fashion so as to make sense of the remarks to which it responds. A flexible interplay between syntactic and semantic features of the linguistic input, and between these and the program's knowledge of the current state and general properties of the world (including its conversational partner), allows for subtleties of comprehension that vastly outrun the capabilities of earlier programs, to say the least.

The program represents a robot named SHRDLU that evinces its understanding partly by verbal responses to its interlocutor, but mainly *by carrying out actions at the human's command*. Access to an internal representation of its own acts (and possible acts) in the world, as well as to a model of the world considered independently, enables SHRDLU to infer the meaning of questions and commands that would otherwise remain opaque.

Winograd's program has at least a partial understanding of "it," and does not assume that "it" must always refer to an object rather than an action. Even within a conversation strictly limited to the discussion of variously colored objects that can be moved and stacked on a table, or put into and taken out of a box, to know what "it" means is to know (to remember or to infer) a good deal about the world and about actors intervening in the world. This is evident from a consideration of the 17 occurrences of "it" in the following dialogue, which indicates the power of comprehension enjoyed by Winograd's program as a result of using its intelligence in interpreting the conversation :

1. Pick up a big red block.

OK.

*{• The system answers "OK" when it carries out a command. In order to pick up the red block, it had to clear it off by finding a space for the green one and moving the green one away. [Figure 1 shows the initial state of the scene.]}*

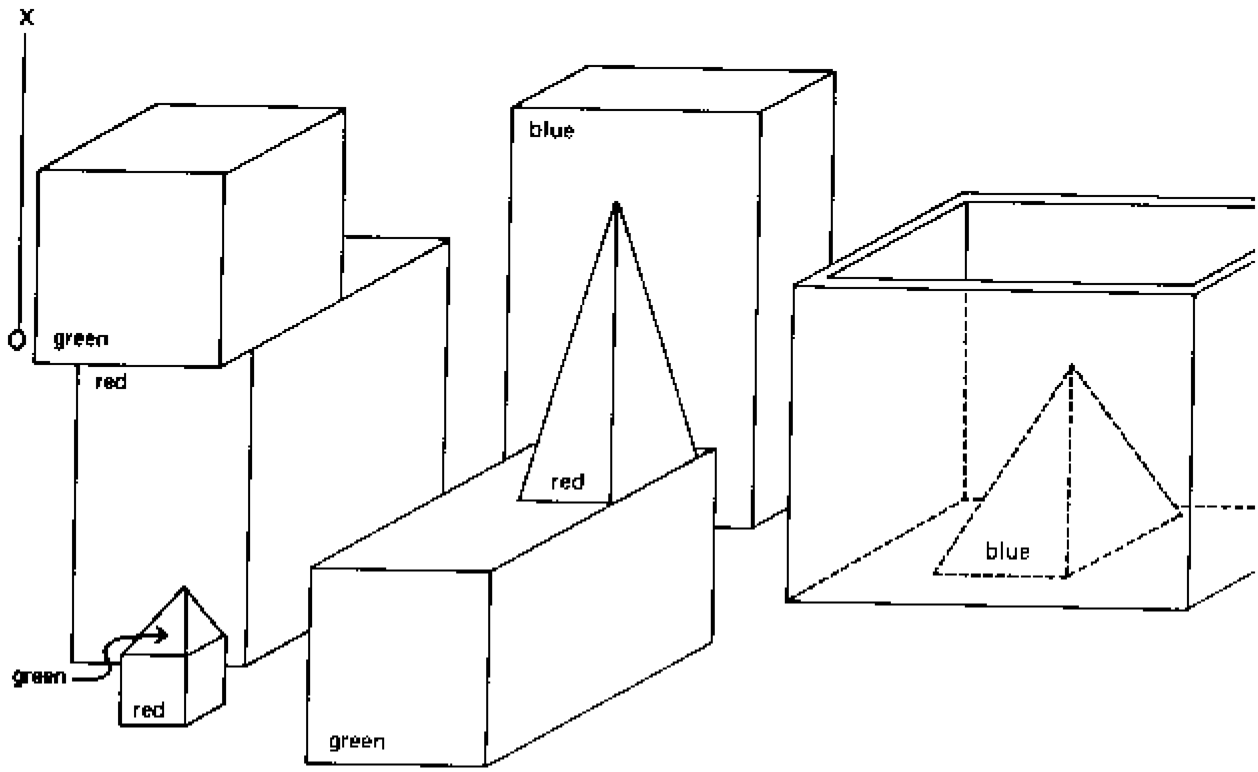
2. grasp the pyramid.

I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.

• There are three pyramids in the scene, so the program does nothing.

3. Find a block which is taller than the one you are holding and put it into the box.

*{• The system has to figure out what is meant by "one" and "it", as well as using deduction to decide which block it is holding.}*

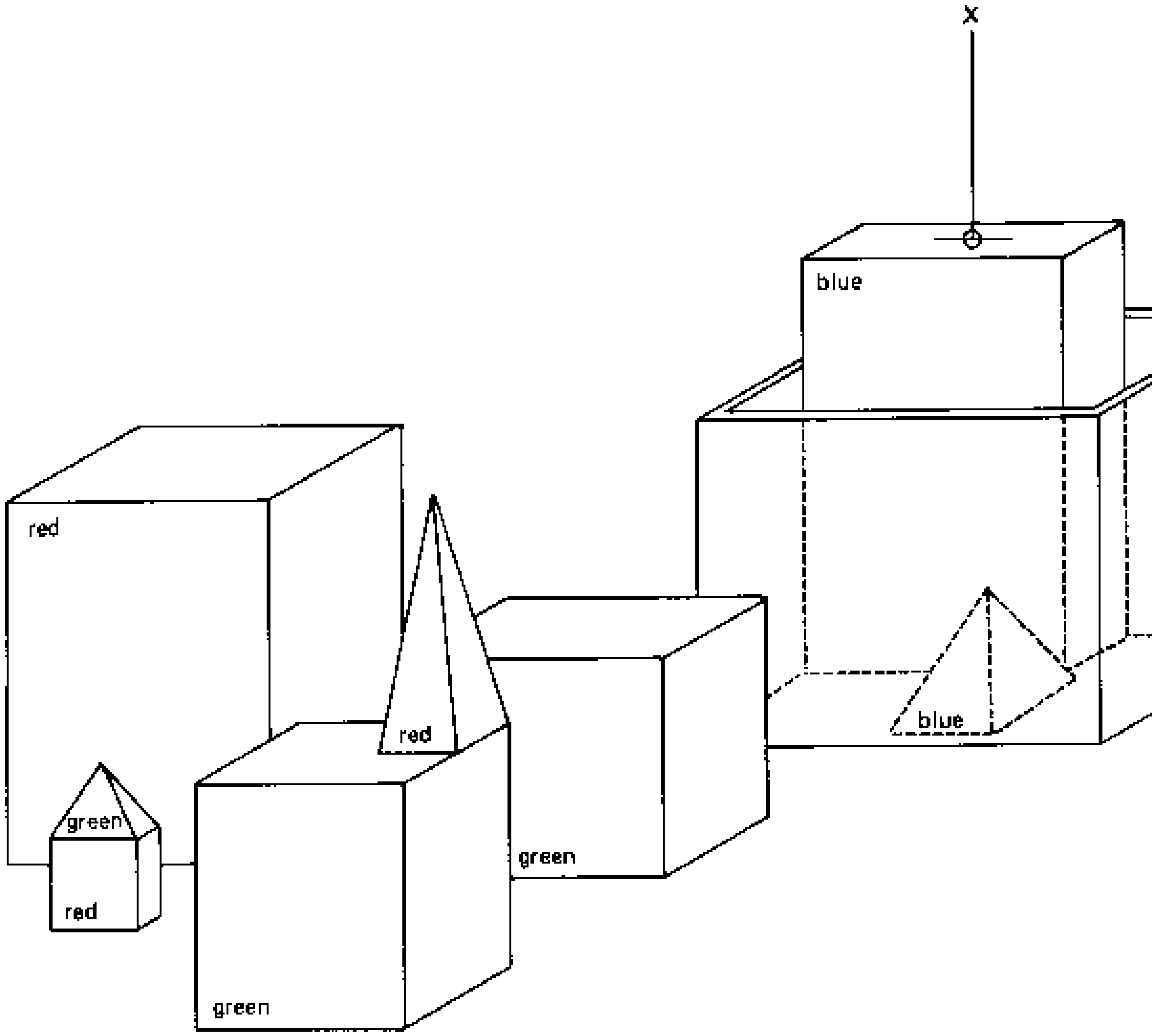



BY "IT," I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.

*{• Printed out to keep us informed.}*

OK.

*{• The robot hand has to set down "the red block" to pick up the taller blue one. Figure 2 shows the result of the system's manipulations so far.}*




4. What does the box contain?

THE BLUE PYRAMID AND THE BLUE BLOCK.

5. What is the pyramid supported by?

*{• Earlier the system responded to "the pyramid" by saying "I don't understand which pyramid you mean." Now, it assumes that we mean the one just mentioned.}*

THE BOX.

6. How many blocks are not in the box?

*{• Negatives such as "not" and "no" are handled whether they appear in clauses or noun groups*

*(such as "no boxes").}*

FOUR OF THEM.

7. Is at least one of them narrower than the one which I told you to pick up?

*{• The phrase "at least one of them" must be analyzed to refer back to the previous question, and "the one which . . ." must also refer to blocks. Noun groups can have complicated numbers such as "at least one," "more than three," "exactly seven," etc. The memory of past events is used to deduce which block is "the one which I told you to pick up." Notice the complicated syntactic structure in which "the one" is actually the object of a clause embedded two levels deeper.}*

YES, THE RED CUBE.

8. Is it supported?

*{• Passives can be used without a "by . . ." phrase.}*

YES, BY THE TABLE.

9. Can the table pick up blocks?

NO.

*{• The semantic rules make this a nonsensical question which therefore must be false.}*

10. Can a pyramid be supported by a block?

YES.

*{• The deductive system finds an actual example, so it knows this is possible.}*

11. Can a pyramid support a pyramid?

I DON'T KNOW.

*{• The question makes sense, but there is no example in the scene, so there is no easy way to tell without being able to examine the robot's programs.}*

12. Stack up two pyramids.

I CAN'T.

*{• The robot tries and fails. We wouldn't even try.}*

13. The blue pyramid is mine.

I UNDERSTAND.

*{• The datum is added to the data base.}*

14. I own blocks which are not red, but I don't own anything which supports a pyramid.

I UNDERSTAND.

*{• Two new theorems are created for proving things about "owning."}*

15. Do I own the box?

NO.

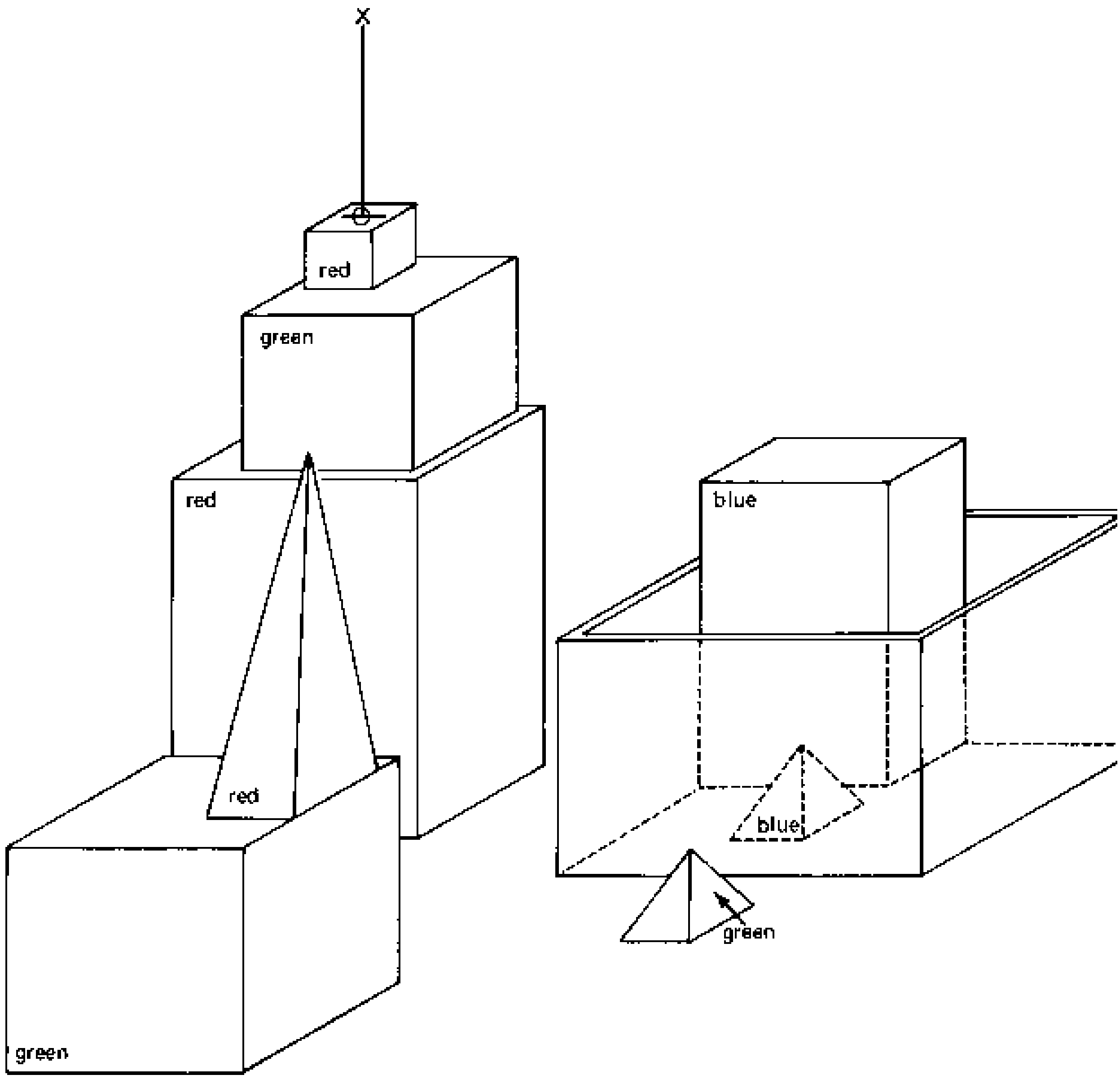
*{• The box supports whatever it contains, including a pyramid.}*

16. Do I own anything in the box?

YES, TWO THINGS: THE BLUE BLOCK AND THE BLUE PYRAMID.

17. Will you please stack up both of the red blocks and either a green cube or a pyramid?

OK.




*{• Logical connectives such as "and," "or," "either," etc. are handled in both the grammar and semantics. A command is executed even if it looks like a question. The command is carried out by the following series of steps. It puts a green cube on the large red block (note that it chooses the green cube with nothing on it), then removes the small pyramid from the little red cube, so that it can use that cube to complete the stack as specified, resulting in the situation of Figure 3.}*

18. Which cube is sitting on the table?

THE LARGE GREEN ONE WHICH SUPPORTS THE RED PYRAMID.

*{• Ordinarily the system would be satisfied with calling this "a big green cube," but since we asked "which cube" it has to be more specific in naming it. The use of substitute nouns (such as "one") and pronouns in answers makes the discourse more natural.}*

19. Is there a large block behind a pyramid?

YES, THREE OF THEM A LARGE RED ONE, A LARGE GREEN CUBE AND THE BLUE ONE.

20. Put a small one onto the green cube which supports a pyramid.

*{• "A small one" must contrast with the earlier phrase "a large block," so this means a small block."}*

OK.

21. Put the littlest pyramid on top of it.

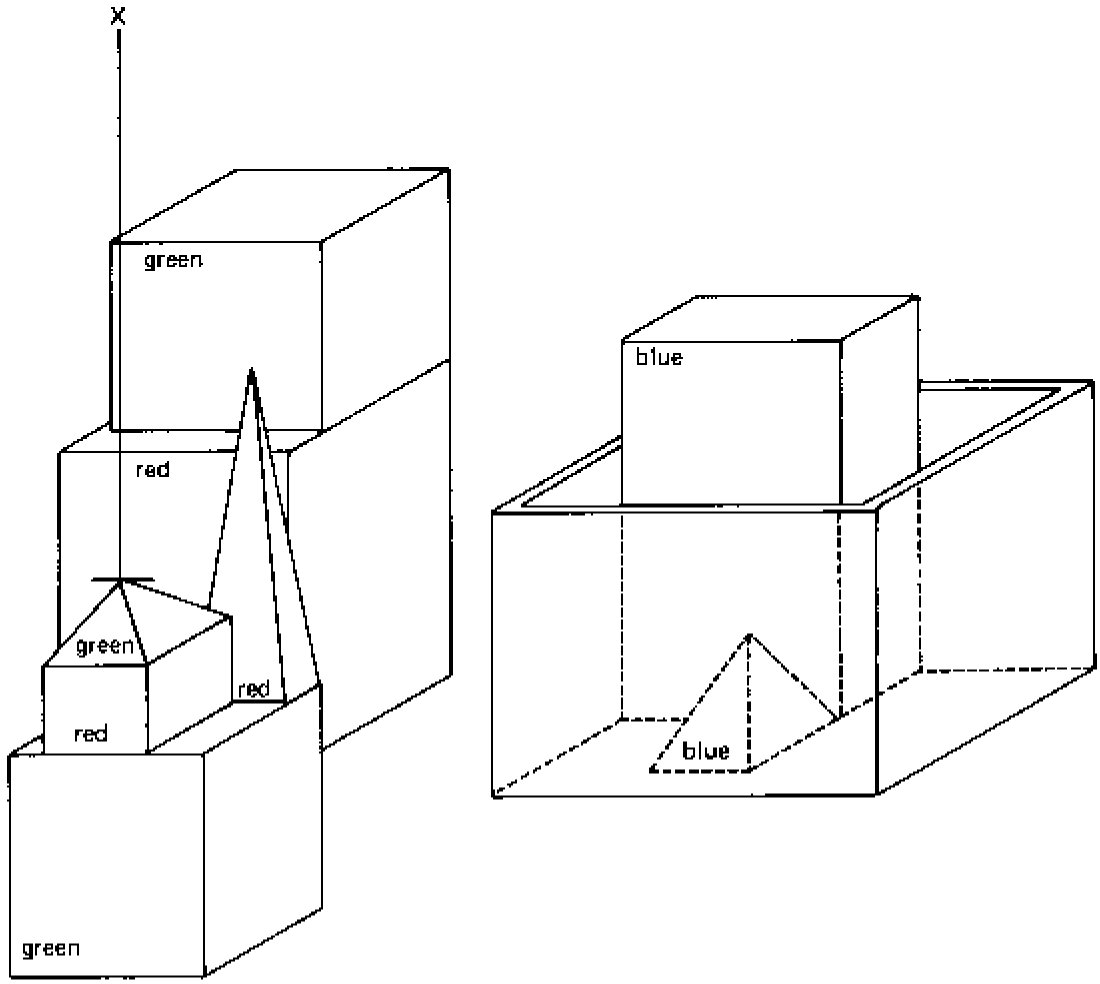
*{• Words like "littlest" are not in the dictionary but are interpreted from the root forms like "little."}*

OK.

22MIT. Does the shortest thing the tallest pyramid's support supports support anything green?

*{• The grammar can handle words like "support" which are both verbs and nouns without getting confused, and can understand possessives. To compute the answer, see Figure 4 for the scene.}*

YES, THE GREEN PYRAMID.




23MIT. What color is it?

BY "IT," I ASSUME YOU MEAN THE SHORTEST THING THE TALLEST PYRAMID'S SUPPORT SUPPORTS.

*{• If we had asked "what side is it?", the system would have assumed that it meant the green pyramid, but nobody would ask the color of something which was just called "the green pyramid," so it had to look farther for the reference of "it".}*

RED.

22. How many things are on top of green cubes?

I'M NOT SURE WHAT YOU MEAN BY "ON TOP OF" IN THE PHRASE "ON TOP OF GREEN CUBES."

DO YOU MEAN

1--DIRECTLY ON THE SURFACE

2--ANYWHERE ON TOP OF?

*{• When there are several interpretations and none is clearly better, the system has to ask the user for more information. The paraphrases were included with the dictionary definition of "on," and the request for more information occurs automatically when the heuristics for resolving ambiguities aren't sufficient. A request like this can also occur when the meaning of a pronoun such as "it or "they" is unclear.}*

2 -- User's reply to SHRDLU's question

THREE OF THEM.

23. Had you touched any pyramid before you put the green one on the little cube?

YES, THE GREEN ONE.

*{• The system deduces from its memory what event is being referred to, and can relate other events to it in time. The analysis includes complex tenses of verb groups, and there are special facilities for keeping track of the various forms of irregular verbs like "have."}*

24. When did you pick it up?

WHILE I WAS STACKING UP THE RED CUBE, A LARGE RED BLOCK AND A LARGE GREEN CUBE.

*{• A time is described by naming the main goal which was being worked on.}*

25. Why?

*{• The system remembers immediately previous sentences to understand questions like this and the following ones.}*

TO GET RID OF IT.

*{• By keeping track of selected parts of the original subgoal tree, the system has some understanding of its own motives.}*

26. Why did you do that?

TO CLEAN OFF THE RED CUBE.

*{• "That" refers to the action mentioned in the answer.}*

27. Why did you clear off that cube?

*{• Here, "that" means the one just referred to.}*

TO PUT IT ON A LARGE GREEN CUBE.

28. Why did you do that?

TO STACK UP THE RED CUBE, A LARGE RED BLOCK AND A LARGE GREEN CUBE.

29. Why did you do that?

BECAUSE YOU ASKED ME TO.

*{• It remembers what it did, not how the request was worded.}*

30. How did you do it?

*{• "It" refers to the action mentioned in the last question, rather than the answer.}*

BY PUTTING A LARGE RED BLOCK ON THE TABLE; THEN LETTING GO OF IT; THEN PUTTING A LARGE GREEN CUBE ON IT; THEN LETTING GO OF THAT CUBE; THEN PUTTING THE RED CUBE ON THAT CUBE; THEN LETTING GO OF THAT CUBE.

*{• This is a description of the top-level subgoals used to achieve the goal.}*

31. How many objects did you touch while you were doing it?

*{• Time references can refer to a time span, using words like "while."}*

FOUR OF THEM.

32. What did the red cube support before you started to clean it off?

THE GREEN PYRAMID.

*{• By remembering what has been moved, it is possible to deduce things about earlier states of the scene. We can use complex nestings of clauses like "before you started to clean it off."}*

33. There were five blocks to the left of the box then.

*{• "Then" must refer to the time mentioned in the last sentence.}*

NO, ONLY FOUR OF THEM: THE RED CUBE, TWO LARGE GREEN CUBES AND A LARGE RED BLOCK.

*{• Earlier in the dialogue, new information about "owning" was accepted at face value. But the system knows that it has complete information about the basic properties of objects, including location. It checks any statements about location or other such properties to see whether it agrees or disagrees.}*

34. Put the blue pyramid on the block in the box.

*{• This is syntactically ambiguous, but is understood without ambiguity since the parsing corresponding to "put the blue pyramid which is on the block in the box" does not make sense in the current scene, while "put the blue pyramid on the block which is on the box" does.}*

OK.

35. Is there anything which is bigger than every pyramid but is not as wide as the thing that supports it?

*{• The system understands quantifiers such as "every," "any," "some," etc., and structures such as "bigger than," "as wide as", etc. Here "it" refers to a noun group within the same sentence but two levels up in the parsing tree.}*

YES, THE BLUE BLOCK.

36. Does a steeple...

SORRY, I DON'T KNOW THE WORD "STEEPLE."

*{• The system interrupts immediately when it sees a word it can't figure out.}*

37. A "steeple" is a stack which contains two green cubes and a pyramid.

*{• The quote marks indicate that we are defining a new word.}*

I UNDERSTAND.

38. Are there any steeples now?

NO.

*{• The new definition is used as a description to recognize steeples.}*

39. Build one.

OK.

*{• The description given in the definition is enough to deduce a plan for building new objects as well as recognizing them. The word "one" means "a steeple", and the plan is carried out. First, one of the large green cubes is cleared off by removing the red cube and two pyramids. Then the two green cubes and the tall red pyramid are stacked up to fit the description.}*

40. Call the biggest block "superblock."

OK.

*{• We can also give objects names.}*

41. Have you picked up superblock since we began?

YES.

42. Why did you drop it?

BY "IT," I ASSUME YOU MEAN THE SUPERBLOCK. I CAN'T EXPLAIN A NONEXISTENT EVENT.

*{• There is no record in memory of any block being dropped.}*

43. Is there anything to the right of the red pyramid?

YES, FOUR THINGS: THE BOX, THE RED CUBE, THE BLUE PYRAMID AND THE BLOCK NAMED SUPERBLOCK.

*{• Once an object has a name, the name is used to help identify it in responses.}*

44. Thank you.

YOU'RE WELCOME!

*{• Idioms can be defined without modifying the grammar.}*

Despite the excessively congratulatory tone of the final comment (which refers to an essentially trivial programming technique well within the range of Weizenbaum's ELIZA, and relied on fairly heavily by K. M. Colby's PARRY), the overall performance of SHRDLU in this man machine dialogue is impressive. So much so, indeed, that the reader's first impressions may blind him/her to a number of serious general limitations that would become more apparent were SHRDLU's human friend to alter the style and topic of conversation. Some of these limitations will be discussed in class, wherein Winograd's program will be contrasted with others yet to be described; but two must be mentioned immediately. The less obvious of these consists in the fact that Winograd's program cannot itself produce sentences of as great a syntactic variety as those that it can accept from its human partner. Winograd's interest is in the interpretation rather than the generation of language, and SHRDLU's capacity to analyze language far outstrips its capability of synthesis. Whereas it can parse grammatical constructions of considerable subtlety, its own speech acts are based on a handful of relatively simple sentence types, which can be transformed in a small number of ways and which are eked out by a few idiomatic expressions such as "OK" and the like. In short, in its generative activities, SHRDLU is considerably more like ELIZA than it appears.

While this undeniably is a limitation, it is however not crippling. Like the human child, or the adult speaker of a foreign language, the program can cope with a high degree of grammatical complexity in the input, and can express itself intelligibly if not stylishly. The matters that SHRDLU is concerned to express (like most matters of human interest) can if necessary be stated by way of a restricted subclass of syntactical forms. That conversation so restricted may eventually strike one as some how "flat," "prosaic," or "dull," is a largely aesthetic rather than a purely linguistic criticism. The complaint that the topic of SHRDLU's conversation is prosaic or dull is, however, more to the point, and is an oblique way of stating the second general limitation of Winograd's program.

SHRDLU would deservedly be rejected as a conversational partner were one offered as alternatives. For there is no question of the program's being able, like humans to talk of many things. From the point of view of Winograd's program, the world consists only of the robot itself; a table top bearing a number of colored objects and a box; and an interlocutor who asks the robot to move the objects about and who plagues it with questions about the world and about its reasons for acting in the way it does. Indeed, there are in fact no objects, no box, no table -- and no robot. The table-top world and the computer's actions in it are simulated by internal data processes paralleled by images on a screen. This visual display is strictly secondary to the workings of the program, and its purpose is merely to aid the human observer wishing to keep track of the "nonlinguistic" computation carried out by the program. The "imaginary" nature of SHRDLU's spatial world does not detract from the specifically linguistic interest of the program, as opposed to its relevance as an exercise in practical robotics, but its relative emptiness does. The vocabulary, environmental information, and general semantics currently available to SHRDLU are greatly impoverished compared with the knowledge mediating one's daily understanding of one's native tongue and SHRDLU's conversation is correspondingly limited. What is even worse than its ignorance of shoes, ships, and sealing wax is its ignorance of the very things it appears to know something about: blocks, pyramids, boxes, and tables. For the same sorts of reason that Colby's neurotic program understands almost nothing of love and hate, even though it can sometimes manipulate these antonymic symbols in an apparently rational manner, SHRDLU understands much less than a person does by the sentence "Pick up a big red block." Winograd's program does not know that blocks are heavy, or that the bigger they are the heavier they are likely to be; it does not know that if they are dropped, or pushed off the edge of the table, they will fall; it does not know that cubical blocks have six square faces, not all of which are visible at the same time. In short, it knows almost nothing about them at all. Even the block-recognizing visual programs to be described know considerably more about blocks than SHRDLU does, and we shall see that their knowledge of cubes and pyramids is pitifully shallow compared with ours.

Nevertheless, it is significant that an extension of Winograd's program could rightly refuse to worry about why the sea is boiling hot. Indeed, SHRDLU's cognitive world has already been extended by the incorporation of the program into a system providing weather information, a system within which the temperature of the sea (and the limits thereof) might well be represented. Similarly, Winograd's program could deduce for itself -- given the taxonomic knowledge that bats are the only flying mammals -- that pigs do not, in fact, have wings. ELIZA, by contrast, would inflexibly parry (joke) the questions by inquiring why s/he felt the temperature of the sea and the anatomy of pigs to be important, and would give no indication of knowing anything about these phenomena. And Colby's cybernetic psychotic, of course, would merely change the subject as soon as possible -- perhaps by associating "pigs" with "police" -- onto topics relevant to its paranoid delusion.

The superior flexibility and common sense thus evident in SHRDLU's conversation (as compared with PARRY, ELIZA, and the early question-answering programs) rely heavily upon two features, features that are likely to be relevant to other projects aiming at artificial intelligence approaching the power of the natural variety. The first is the "heterarchical" organization of the various components of the overall program, while the second is the representation of different types of knowledge on which the program can draw in interpreting and replying to the comments of its human friend.