Natural Language Understanding

Language is more than words; "meaning" depends on context, and "understanding" requires a vast body of knowledge about the world

Teaching computers how to deal with natural languages—English or French as opposed to FORTRAN—has been a major issue in artificial intelligence research (AI) ever since the discipline began nearly 30 years ago. Quite aside from the fact that literate computers would be far less intimidating to lay users than the current variety, language appears to be central to the workings of the human mind, and thus to AI's attempts to understand the mind. Some anthropologists have even suggested that it was the acquisition of language that marked the evolutionary transition to modern man some 50,000 years ago.

In any case, AI is based on the premise that the mind can be modeled as a processor of symbols—in essence, as a computer program. And language is the quintessential symbolic processing problem.

In 1949, when computers were still very new and everything seemed possible, Warren Weaver, director of natural science for the Rockefeller Foundation, distributed a memorandum to about 200 friends outlining a proposal for "the solution of worldwide translation problems." Computers had been used to break codes during the war, he pointed out. So why not translation? "When I look at an article in Russian," he wrote, "I say, *This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.*"

Weaver's idea was to create a kind of automated bilingual dictionary: the machine would translate each word of the input text into an equivalent word in the output language, and then rearrange the result to fit the output language's word order. Obviously the process would not be quite that simple—some words have different meanings in different contexts, for example, and every language has idioms that make no sense when they are translated word for word—but on the whole it seemed to be mainly a problem of vocabulary.

Weaver's colleagues leapt to the challenge. Research groups were formed both here and abroad. Programs were written, conferences were held, and government funding was provided. In 1954 the journal *MT* (for Machine Translation) was founded. Glowing reports began to appear in the press, just as similar reports are now appearing about current AI work (*Science*, 24 February, p. 802).

And yet, as time went by a certain uneasiness began to creep in. Legend has it that one early researcher asked his computer to translate "the spirit is willing but the flesh is weak," first into Russian and then back into English. The result: "The vodka is good but the meat is rotten."

Actually, the real programs were not even that clever. The first Russian/English translation program was written by A. G. Oettinger in the mid-1950's. A sample of the output, with the possible interpretations of each word listed in parentheses: "(In, At, Into, To, For, On) (last, latter, new, latest, lowest, worst) (time, tense) for analysis and synthesis relay-contact electrical (circuit, diagram, scheme) parallel- (series, successive, consecutive, consistent) (connection, junction, combination) (with, from) (success, luck) (to be utilize, to be take advantage of) apparatus Boolean algebra."

> The problem is to figure out what it means to "understand."

Translation of the translation: "In recent times Boolean algebra has been successfully employed in the analysis of relay networks of the series-parallel type."

But the really frustrating thing was that the programs did not seem to get any better. By 1966, when Oettinger's program was still pretty much the state of the art and no improvement was in prospect, the uneasiness had matured into a widespread sense of futility; that was the year the National Research Council's Automatic Language Processing Advisory Committee recommended that most of the funding for machine translation research be terminated.

The fundamental problem was clear. Whatever it was that language encoded, it was not just a matter of words and definitions and vocabulary. Somewhere behind the surface structure of human language there lay an enormous body of shared knowledge about the world, an acute sensitivity to nuance and context, an intuitive insight into human goals and beliefs. Any machine that was going to translate between languages would first have to "understand" what was being said—which meant that it would somehow have to "know" a very great deal about the world beforehand. As Israeli researcher Yehoshua Bar-Hillel had written in 1960, "A translation machine should not only be supplied with a dictionary but a universal encyclopedia."

And so, machine translation died. But if it was a failure, at least in that early incarnation, it did lay the foundations for the study of language in the closely related field of AI. In fact, the Research Council concluded in that same 1966 report that computational linguistics remained worthwhile as a scientific endeavor, and that funding should be continued for the effort to write programs able to understand language.

The problem, of course, has been to figure out what it means to "understand."

From an operational point of view, for example; one could say that a program "understands" a phrase when it makes the appropriate response. Thus AI's early efforts at template matching: the computer simply searched for key words in stereotyped sentence structures, and then gave back an equally stereotyped response. This equals "understanding" in roughly the same sense that the family dog understands words like *dinner* or *walk*. But in restricted settings it has been surprisingly effective. The most famous program of this type, ELIZA, written in 1966 by Joseph Weizenbaum of the Massachusetts Institute of Technology, was so good at imitating a nondirective psychotherapist that people would quickly find themselves sharing intimate details of their lives with the computer. (Subject: "I'm unhappy." ELIZA: "Do you think coming here will help you not to be unhappy?" . . .) Template match-

ing is still widely used in commercial natural language systems.

However, a substantial minority of AI researchers has always felt unsatisfied with the stimulus-response approach to understanding. They want their programs to work the same way humans work, to somehow be models of the mind. This is sometimes called the "scientific" approach to AI, as opposed to the "engineering" approach; the computer is simply seen as a tool for testing the models, much as physicists embody their ideas in equations.

One such model, developed in the early 1970's by Stanford University's Terry Winograd for his doctorate at Massachusetts Institute of Technology, was that understanding a query or a command is equivalent to constructing a program to produce the appropriate response. To embody this concept Winograd wrote SHRDLU, a system that converted words into program fragments, and sentence structure into an ordering for the fragments. SHRDLU also worked remarkably well. In fact it was the first effort to deal with syntax, semantics, and reasoning ability in a completely integrated way.

On the other hand, SHRDLU only had to deal with a very narrow universe: its "blocks world" consisted of a simulated robot arm manipulating simulated blocks on a simulated tabletop. Extending it to wider domains meant sacrificing more and more of its power. In addition, SHRDLU had trouble with such elementary English words as *the* and *and*. So in retrospect it was a lot better than template matching, but somewhat less fluent than a human 4 year old.

In the effort to do better, a key issue for many researchers has been the code of language itself, the grammar. One thing made painfully clear by the machine translation experience, and reenforced by modern linguistics, is that this code is not simple. Not only can different sentences mean the same thing—*Bill bought the car from Fred* equals *Fred sold the car to Bill*—but meaning itself depends critically on context. Consider a sentence such as *The duck is ready to eat*. Consider the verb *made* in the sentences *Sue made the bed*, and *Sue made an A on the test*.

In practical terms this means that a natural language program should be able to "delinearize" a text, deciphering how the words and sentences relate to each other so that it can represent their meaning in some deeper grammatical structure. In high school this is called parsing, or diagraming, a sentence; AI researchers have in fact devised some very powerful parsing engines over the years, most notably the Augmented Transition Networks (ATN's) first developed in the early 1970's by William Woods and his colleagues at Bolt Beranek and Newman, Inc., in Cambridge, Massachusetts.

In theoretical terms, however, there is the intriguing possibility that this deep grammatical structure is somehow universal, that there exists a unified theory of all possible human languages—and that when this theory is finally found, we will have discovered something profound about the human mind.

This is hardly a new idea. Weaver postulated a universal "interlingua" for machine translation in his 1949 memorandum, for example. And it has often been suggested that children learn language so rapidly because the ability to do so is somehow hardwired into their brains.

Since the late 1950's, however, the universality idea has most often been associated with MIT linguist Noam Chomsky, who popularized it in connection with his theories of "transformational" grammar. These theories, which involve formal, quasi-mathematical rules for manipulating such factors as word order, tenses, and word endings, brought an unprecedented rigor and precision to linguistics and have been highly influential. However, transformational grammar has often seemed at odds with what psycholinguists have learned about the way humans go about language. Moreover, the formalism has also proved to

Say What?

Most of AI's natural language effort has focused on the written word, simply to avoid the extra complexity of deciphering sound waves. But human beings do a great deal of their communicating via the spoken word, and in many settings—say in a spacecraft where the pilot has his or her hands full with the controls—a verbal command would certainly be the fastest and most natural way to communicate with a computer.

Actually, systems that can recognize individual words have been commercially available for some time now. There are video games that can do it for less than $100. But even the best of them is simply matching acoustic signals against a set of electronic templates. Not only does the speaker have to enunciate each word separately and distinctly—a very unnatural way of talking—but each new user first has to "train" the machine to his or her voice by pronouncing every word of the vocabulary beforehand; vocabularies of more than a few hundred words are thus impractical.

It would be nice if computers could understand continuous speech; unfortunately, continuous speech poses a horrendous recognition problem. The acoustic signal of a given word varies according to where it is in the sentence, what words surround it, and what the sentence is saying. Even finding the boundary between two words can be difficult—consider the phrase "gas station."

The first major, coordinated effort at speech understanding was sponsored by the Defense Advanced Research Projects Agency (DARPA) between 1971 and 1976. The goals were relatively modest—a 1000-word vocabulary, for example, a 10 percent error rate, slower than real-time processing—but the program did result in several demonstration systems. More importantly, it sparked the first attempts to integrate "front-end" approaches, such as phonetics and signal processing, with "back-end" approaches, in which the computer attempts to sort out the acoustic ambiguities using knowledge about the speaker, the situation, and the structure of the language.

Since the DARPA study, research on continuous speech recognition has been largely dormant in this country; the sole exception is the ongoing effort at IBM's Thomas J. Watson Laboratories in Yorktown Heights, New York, where the goal is a real-time office dictation machine with a vocabulary of some 5000 words. In Japan, however, the so-called "Fifth Generation" project has announced that one of its long-range goals will be a 10,000 word, speech-activated typewriter with the ability to understand hundreds of different speakers. Some observers in the United States believe that, with a substantial application of resources, a limited version of such a system could be available in the 1990's.—**M.M.W.**

be remarkably cumbersome in practical computer programs; the whole thing has thus acquired a bad odor in certain seg ments of the AI community.

On the other hand, not everyone has given up. There have been a number of attempts in recent years to formulate non-Chomskian grammatical theories that are both psychologically realistic and computable. One notable example is the Lexical Functional Grammar (LFG) developed by Ronald Kaplan and Joan Bresnan at the Xerox Palo Alto Research Center. LFG deals with functional roles—subject, predicate, and so forth—that have nothing necessarily to do with word order or phrase structure. Its formal representations of language tend to be intuitive, elegant, and mathematically tractable; moreover, the Japanese have testified to its universality by designating LFG as a prime candidate for the grammatical formalism of their Fifth Generation project (*Science*, 24 February, p. 802).

Meanwhile, a very different approach to deep structure has been taken by those researchers who argue that grammar, per se, is beside the point. The real business of human language understanding is not going on at the level of language itself, they maintain. It happens well below that at the level of "concepts."

A notable proponent of this point of view is Roger Schank of Yale University. In his model of "conceptual dependency," under development since the early 1970's, sentences are mapped into elaborate data structures organized around a handful of "semantic primitives." For example, verbs such as *move, walk*, or *lift*, which involve changing the physical location of an object, are all mapped into a single primitive ACT called "PTRANS." An attribute such as *Mary is dead* maps into a primitive STATE, *Mary HEALTH(-10)*, and so forth.

The upshot of all this, according to Schank, is a formalism that is both psychologically plausible and independent of the particular language in question. And indeed, as implemented in the program MARGIE in the mid-1970's, Schank's scheme was quite impressive. MARGIE was able to make inferences from input sentences (*John hit Mary* implies, among other things, *Mary might hit John back*). And it was able to paraphrase sentences (*John said he was sorry* might become *John apologized.*) Moreover, since the ability to paraphrase a sentence is only a short step away from the ability to translate that sentence to another language, MARGIE and its relatives have been used for machine translation—although fluent translation remains as elusive as ever.

It must be said, however, that the system as described leaves a bit to be desired. *Kiss*, for example, is represented as *MOVE lips to lips*; the "understanding" extends only to the physical world. As Weizenbaum pointed out in his 1976 book *Computer Power and Human Reason*: "It may be possible, following Schank's procedures, to construct a conceptual structure that corresponds to the meaning of the sentence, 'Will you come to dinner with me this evening?' But it is hard to see—and I know this is not an impossibility argument—how Schank-like schemes could possibly understand that same sentence to mean a shy young man's desperate longing for love."

On the other hand, a less pessimistic observer might say that the solution to the problem is obvious: give the program more knowledge. In particular, give it more knowledge about human nature and human social interactions.

An enormous amount of mutual accommodation and inference goes on, much of it below the level of the literal meaning of the words.

In fact, this has become a major goal of AI natural language research in recent years. Schank and his colleagues, for example, have broadened their formalism to take account of such things as purpose and expectation. There are *goals* and *plans*; there are *scripts*, which list the typical sequence of actions one might expect in, say, a restaurant; and there are *themes*, which allow the computer to draw inferences about a person's goals from, say, his occupation (lawyer) or his relationship with others (love).

A closely related activity is the study of discourse, the communication between several individuals. "The AI approach [to language] has been very limited," says Barbara J. Grosz of SRI International. "It assumes that there is just one agent involved, which knows everything it needs to know. Well, that kind of model breaks down in obvious ways." In a conversation—be it an everyday chat over the telephone or dialogue in a story—the participants may differ in what they themselves know or want, and they may very well differ in what they believe of each other. An enormous amount of mutual accommodation and inference goes on, much of it below the level of the literal meaning of the words.

She points out, for example, that a sentence such as *Can you pass the salt?* is not just a request for true-false information about the world. A robot who replied *yes* to such a question would be very annoying. The sentence is in fact an attempt to elicit action; the listener—or the computer—has to understand that.

The striking thing about this line of research is how quickly AI starts to sound like a kind of applied philosophy. In fact, much of the work takes inspiration from the "speech acts" theory of the University of California, Berkeley, philosopher John Searle. Essentially, Searle analyzes language in terms of commitments between speaker and hearer; even a supposedly objective statement such as *It's 3 o'clock* is seen as a commitment by the speaker that the proposition is true.

Another example of this straining at the boundaries is the Center for the Study of Language and Information (CSLI), founded at Stanford in late 1983. Among the participants are Grosz, Winograd, Kaplan, Bresnan, and more than a dozen other researchers in the Palo Alto area; in fact, CSLI is probably the first real attempt to maintain a dialogue among *all* the language-related fields, including linguistics, philosophy, AI, computer science, and psychology. The enthusiasm there is palpable, albeit the center is still in its infancy; people maintain that they never realized how much they had to say to each other.

No one, of course, knows where natural language research will lead. The most ambitious hope is that, through some magic of Lexical Functional Grammar, or conceptual dependency, or "speech acts," or the cross-fertilization at CSLI, scientists will one day come to understand what "understanding" is. More realistically, they may simply learn how to differentiate those areas of human intellect that can be modeled by symbol processing from those that require some other model. Or perhaps they will only succeed in designing much better ways to communicate with their computers.

In any case, the goal seems worth the effort.—**M. MITCHELL WALDROP**

This is the third in a series of occasional articles about artificial intelligence. Previous articles appeared in Science, *24 February, p. 802, and 23 March, p. 1279.*