Associations or Rules in Learning Language?

Linguists and computer scientists debate whether so simple a matter as learning to form the past tense of verbs requires an understanding of the structure of language or whether the past tense can be learned without rules

LAST year, David Rumelhart of the University of California in San Diego and James McClelland of Carnegie Mellon University suggested that children learn language not by learning rules of grammar and word composition, but by learning analogies. In other words, they do not subconsciously grasp concepts such as "noun," "verb," "prefix," and "root," but instead reason by deciding that "this word sounds like that word."

As evidence that children learn language by analogy, Rumelhart and McClelland produced a computer model that "learns" English without ever knowing any rules and that makes some of the same mistakes as it learns that children do.

Now that hypothesis is under attack by Steven Pinker of the Massachusetts Institute of Technology and Alan Prince of Brandeis University who say that the computer system does not mimic human learning and who contend that children really do learn rules when they learn a language.

The language debate is actually an old one in a new guise. It is, says Prince, "a classical issue that keeps popping up in studies of philosophy and psychology whenever people try to understand what it is that underlies cognitive behavior and regulates thought." More than two centuries ago, for example, classical empiricists, led by philosophers George Berkeley and David Hume argued for associations as a basis for knowledge whereas rationalists, led by philosopher René Descartes argued for rules.

The argument over rules versus analogies touches on fundamental issues in computer science and linguistics. It is part of a larger dispute within the artificial intelligence community over the meaning of a new form of ruleless computer systems and it bears on the question of the very nature of language. As such, it is attracting an unusual amount of attention.

Pinker explains the importance of the model to the artificial intelligence community as follows. "Language acquisition," he says, "is the jewel in the crown of cognition. It is what everyone wants to explain." The reason, he says, is that "language acquisition is the toughest nut to crack. In a sense, language acquisition defines what it is to be an intelligent human being." And if a ruleless computer model is sophisticated enough to account for language acquisition, Pinker says, "it should be powerful enough to handle other aspects of cognition."

A few years ago, artificial intelligence researchers discovered that they could make computer programs that "learned" without ever being given symbols and rules to manipulate the symbols. This approach, which was a radical departure from traditional artificial intelligence research, has sharply divided the scientific community. It is the basis of the language acquisition model proposed by Rumelhart and McClelland.

"Language acquisition is the jewel in the crown of cognition. It is what everyone wants to explain."

The ruleless systems are called connectionist networks because they consist of densely connected network of processors. The processors transmit signals that vary in intensity according to the strengths of the signals that each unit of the network receives. Connectionist models work by learning routines to adjust the strengths of the connections between the processors. No rules are fed into the system but, at the end, when the network adjusts itself, something very much like rules are learned.

Rumelhart and McClelland wanted to explain with their connectionist model what really goes on when children learn to speak. Are there rules that children gradually discover or do they learn by a process that more closely resembles forming analogies? And if language acquisition can be understood, can it also be modeled by a computer? In particular, the two researchers wanted to explain the strange process that children go through when they learn to form the past tense of verbs.

The past tense in English is almost incredibly simple. To form the past tense of the verb, you almost always add "ed" to the end of it, unless it is an irregular verb, such as "go" or "bring." Among the 150 or so irregular verbs, "go" is in a class by itself. The rest of the verbs fall into about 20 groups, such as the group containing the verbs keep, sleep, and weep. All the verbs in a group form past tenses in the same way.

Because the past tense is so simple, says Prince, "linguists have not studied it much. When you go to graduate school, it is not the sort of thing you linger over."

But psycholinguists have discovered that when young children learn to speak, they start out by forming irregular past tenses correctly and then they get worse—they overregularize. Finally they learn the correct forms again. For example, children start out by saying "brought" and "went." Then they switch to "bringed" and "goed" before they relearn the correct irregular forms.

The standard explanation of overregularization is that children when they first learn to speak, memorize words one by one without regard for any relations between them. Later, they discover the past tense rule and run amok with it, overregularizing, because they do not grasp the structure of the language. Finally, they learn the exceptions to the past tense rule and their speech becomes correct again. The idea is that children eventually learn the past tense rule for regular verbs and learn the irregular past tenses by analogy.

Rumelhart and McClelland started out by assuming that this standard explanation is correct. Only after they developed their connectionist model of language acquisition did they question it. Rumelhart, in fact, used to illustrate the observation that children learn rules for forming the past tense by playing for his students a tape of his own little boy, who was 5 years old when the tape was made.

In the recording, Rumelhart asked his son what grade comes before the seventh grade. "Sixth," the boy replied. Then Rumelhart asked what grade is before the sixth grade. "Fifth," the boy said. What is before fifth? "Fourth." What is before fourth? "Thirdth." What is before third? "Secondth." What is before that? "Firsth."

Then Rumelhart asked the question in the opposite order. What grade is after kindergarten? His son replied "First." What is after first? "Second." Rumelhart continued up to grade seven and, this time, the boy got all the words right.

"I would play this tape for students and

Language guidelines

Children learn language in a social context, but the question is, What rules, if any, do they learn to follow?

would tell them that it was obvious that the kid had learned a general rule," Rumelhart says. He did not worry about the fact that his child got the words wrong when he went in descending order and got them right when he went in ascending order.

Rumelhart told his students that the past tense is learned in the same way—with rules. Then Rumelhart and McClelland noticed that connectionist networks tend to overregularize in the same way that children do when they learn to speak. Rumelhart explains that "when there are some things with regular patterns and others with unusual patterns, the networks learn the regular patterns first and apply them where the more unusual ones should be applied." Only later does the network learn the unusual patterns.

So Rumelhart and McClelland "built a very simple-minded connectionist model, mainly to show how overregularization occurs," Rumelhart says. "To our surprise, not only did overregularization of the past tense occur, but the patterns were strikingly like those that occur in children's speech." This finding, Rumelhart emphasizes, "was as much a surprise to us as to anyone else."

The two researchers decided that perhaps the model was coming closer to human language acquisition than they had dreamed possible. "Of course, we recognized that it is an oversimplification in numerous ways. For example, it only has a 400-word vocabulary. But it turned out that, as far as we could tell, our little demonstration looked to be as good an account of language acquisition or better than anyone had come up with," says Rumelhart. "So we tried to understand wherein lies the power."

To teach the computer the past tense, Rumelhart and McClelland first gave the computer a small set of predominantly irregular verbs with their past tenses. The computer correctly "learned" the regular and irregular forms. Then they gave the computer a much larger list of predominantly regular verbs, whereupon the computer began overregularizing. Finally, the computer sorted out the regular and irregular forms and gave mostly correct past tenses.

Rumelhart and McClelland decided that the computer, which learns by analogy, provides a model for how the human brain learns. "I really believe that being able to give correct responses to novel situations is more like using an analogy than a rule," Rumelhart says. "But the problem with the analogy hypothesis was that no one knew how to apply it. How could you learn language by using analogies? The connectionist model more or less does this automatically. It is hard to stop it from doing it."

Pinker and Prince decided to take a closer look at Rumelhart and McClelland's model and, after a detailed analysis, concluded that it does not accurately model language acquisition and that, in fact, children learn rules rather than analogies when they learn the past tenses of regular verbs.

"In many ways, Rumelhart and McClelland's model is very impressive," Pinker notes. "It accounts for rulelike behavior without positing any rules. It is considered a triumph of connectionism." And connectionism, Pinker notes, is frequently hailed as a promising approach to understanding the workings of the human brain. The densely connected computer circuits are said by neuroscientists to have at least a vague resemblance to neural nets and the fact that the connectionist models learn without rules has been seized on by psychologists who point out, according to Pinker, that "many aspects of cognition are not crisp and rulelike."

One of Pinker and Prince's criticisms is that the overregularization of past tenses in the computer model is caused by a statistical trick that does not resemble what happens when children learn to speak. The computer at first is presented with just ten verbs, eight of which are irregular. At this point the computer "learns" the irregular past tenses correctly. Then the computer is flooded with several hundred verbs, 80% of which are regular. "Now, because the model is overwhelmed with regular verbs, it starts behaving like a child," Pinker says. But the reason the computer starts overregularizing is that "the world in which the model lives changes radically," Pinker notes. "It goes from 80% irregular verbs to 80% regular. Basically, it's a trick and it's based on a changed input."

So Pinker and Prince looked at children and asked whether their vocabulary changes as they grow older so that more and more verbs in their vocabulary become regular. They found, says Pinker, that this is "completely untrue. The proportion of regular to irregular verbs in children's vocabularies stays at 50–50 over a period of several years. But during this time, the children start to overregularize. The explanation lies not in the environment but in the head of the child."

Rumelhart says that he and McClelland will formally reply to Pinker and Prince's criticisms, but that it will take them some time. Rumelhart and McClelland agree with some of the criticisms of their model but disagree with others. But the model is promising enough, says Rumelhart, that "we would like to do it over again and do it better."

In the meantime, the question of how children learn language is still open to debate. If it is strictly by analogy, then why are there such usages as "righted the boat?" If a noun, like "right" is made into a verb, it cannot produce a new verb root, just a new verb, Pinker and Prince point out. So reasoning by analogy, which would produce the phrase "rote the boat," does not occur.

On the other hand, if regular forms in English are learned strictly by rule, then why did Rumelhart's son say "thirdth" when he was going through the grades in descending order and say "third" when he was going in ascending order? It evades the question to say that the truth lies somewhere in the middle. Instead, there is still no consensus on how children learn even the simple-sounding past tense rule. ■ **GINA KOLATA**

ADDITIONAL READING

D. E. Rumelhart and J. L. McClelland, "Learning the past tenses of English verbs: Implicit rules or parallel distributed processing?" in *Mechanisms of Language Acquisition*, B. MacWhinney, Ed. (Erlbaum, Hillsdale, NJ, 1987).

S. Pinker and A. Prince, "On langauge and connectionism: Analysis of a parallel distributed processing model of language acquisition," occasional paper #33 (Massachusetts Institute of Technology, Cambridge, MA, 1987).