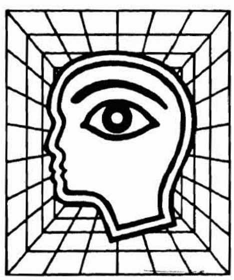
6

Language and Communication



The notion of a shared model is inherent in the word "communicate," which is derived from the Latin *communicare*, to make common. People communicate to command, question, inform, promise, threaten, amuse, arouse, and convince other people. Thus a person has an intended idea, request, or command to communicate with another. The idea, request, or command encoded into their shared language is communicated to the recipient, who derives the meaning of the message using a "model" of the person communicating the message, the context of the communication, the appropriate "world knowledge," and knowledge of the language.

A language is a set of vocal or written signs and symbols that permits a social group to communicate, and facilitates the thinking and actions of individuals. Civilized life in its present form would probably be impossible without the use of spoken and written language. Language, in the full sense of the term, is species-specific to man. Members of the animal kingdom have the ability to communicate through vocal signs, facial expressions, and by other means, but the most important single feature characterizing human language is that people are essentially unrestricted in what they can talk about.[9] As will be described later, animal communication

[9]We say "essentially unrestricted" because, while poetry and creative writing make an attempt, there is no adequate way to describe sounds, smells, taste, and other experiences in a written or spoken languages.

systems are, by contrast, very tightly circumscribed.

Human thought and language are closely linked; as a matter of fact, some believe that the language we speak critically influences both the way we think and the way in which we perceive reality (the Sapir-Whorf hypothesis).

This chapter describes the nature of communication using language, how language encodes meaning, and work in computational linguistics that attempts to provide a basis for computer *understanding* of natural language. We will find that building computer systems for effective interaction with people requires that language be considered in the context of a communication situation. In this larger context, the relationship between participants in a conversation, and their states of mind, are as important to the interpretation of an utterance as the linguistic components from which the utterance is formed. We will discuss a number of questions that are still being actively, and sometimes heatedly, debated:

- Can animals, particularly chimps and gorillas, acquire and creatively use natural language?
- Must children be trained to acquire language, e.g., by their parents, or do children have an innate capability to form a "theory" of language on their own?
- Does the language one speaks determine the way one perceives the world, or are people's world views independent of the language they speak?
- What is the purpose of communication, and to what extent is communication possible without language?
- How is the human brain organized to provide linguistic competence?
- Are there things you can express in language that cannot be expressed in any other form of communication?
- What is the relationship of language to reasoning and intelligence? Can a person be intelligent without some form of language (e.g., spoken, written, or sign language)?
- What are the limits of a computer's ability to employ natural language—are there linguistic expressions whose meanings cannot be derived by a machine, and if so, what is their general nature?
- Is it possible for a machine to truly understand natural language, or is the machine, at best, merely manipulating tokens so that it only appears that understanding is taking place?

The first part of this chapter concerns human and animal communication, and the second part, machine communication.

LANGUAGE IN ANIMALS AND MAN

If we define language broadly enough, then it can be said that both animals and humans are capable of communicating with other members of their species via language. However, in animal communication the language is very limited, restricted to a number of sounds associated with signaling danger, establishing territory, indicating anger, etc., without the *creative* aspect of human language in which a set of basic sounds is used to express indefinitely many thoughts, and respond appropriately to an indefinite range of new situations.

Brain Structures Associated with Language Production and Understanding

What little is known about the role of the brain in communication has been derived by studying the relation of brain damage to performance [Geschwind 79]. Figure 6-1 shows the regions of the human brain that have been identified as being relevant to linguistic activity.

Broca's area is named after Paul Broca, who in the 1860s noted that damage to a particular region of the cortex on the side of the frontal lobes gives rise to speech disorder (aphasia). He showed that damage to this area on the left side of the brain causes aphasia, but damage to the corresponding area on the right side leaves speech intact. In 1874, Karl Wernicke identified an area on the temporal lobe of the left hemisphere that plays a crucial role in communication. By relating defects in the Broca and Wernicke areas to loss of performance, Wernicke formulated a model of language production.

In this model, the underlying "structure" of an utterance arises in Wernicke's area and is transmitted to Broca's area through a bundle of nerve fibers called the *arcuate fasciculus*. Broca's area develops a "program" for vocalization that is then passed to the face area of the motor cortex, to activate the appropriate muscles of the mouth, lips, tongue, and larynx. When a word is heard, the sound is received by the auditory cortex and then passed to Wernicke's area where it is "understood." When a word is read,

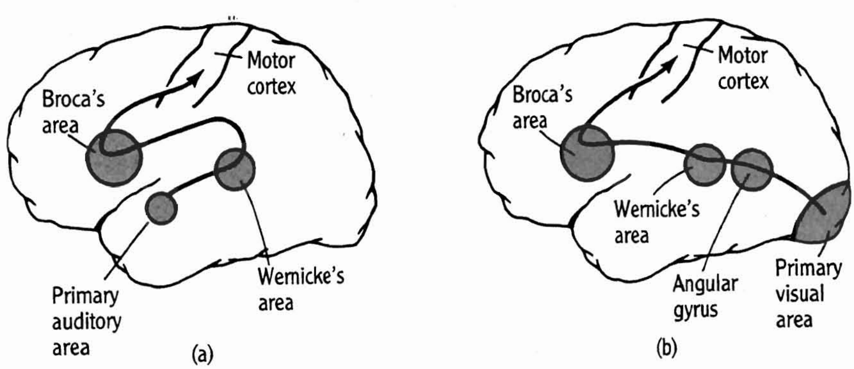


FIGURE 6-1 Brain Signal Flow for Language Production and Understanding.

(a) Speaking a heard word. (b) Speaking a written word. (After N. Geschwind. *Sci Am* 243:180-199, 1979.)

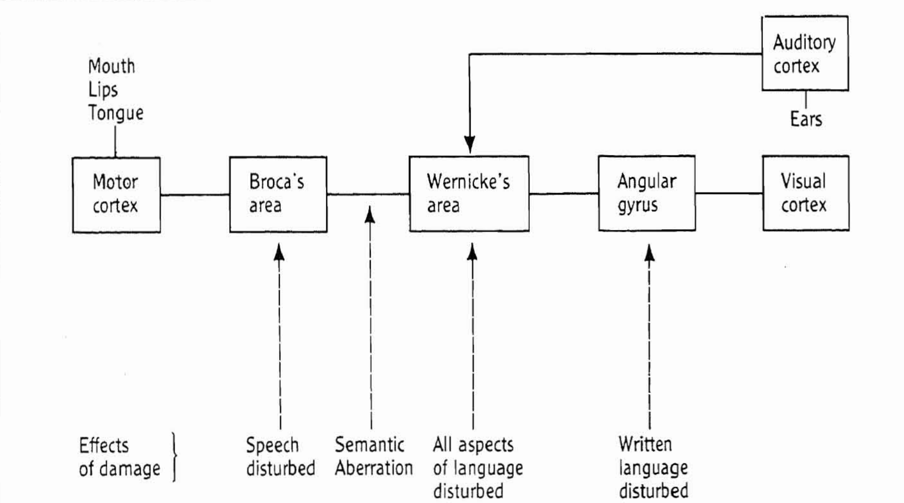


FIGURE 6-2
Schematic of the Structures Involved in Language Production and Understanding, Showing Effects of Brain Damage.

the information from the visual cortex is transmitted to the *angular gyrus* where it appears to be transformed so as to be compatible with the "auditory form" of the word; it is then transmitted to Wernicke's area.[10]

Figure 6-2 shows the effect of damage to each of the components of the brain's linguistic communication system: When Broca's area is damaged, speech is no longer fluent or well articulated. When the path from Wernicke's area to Broca's area is damaged, semantically aberrant speech is produced, but if Wernicke's area is intact there will be normal comprehension of spoken and written communication.

[10]It is difficult to understand what evolutionary mechanism could have prepared the angular gyrus for this role! Some relevant ideas can be found in Box 2-3.

Damage to Wernicke's area disrupts all aspects of the use of language, as can be seen from the central role that it plays in the system shown in Fig. 6-2. Finally, damage to the angular gyrus disrupts the signals from the visual cortex to Wernicke's area and causes difficulties in dealing with written language.

This view, that there are discrete cerebral centers performing specific aspects of language processing, has been called the *localizationist-connectionist* view. As discussed by Springer and Deutsch [Springer 85], present-day investigators with more holistic views of brain function contend that the situation is more dynamic than implied by localization theories, involving simultaneous interactions of many areas for each brain function.

Even to the extent that specific regions of the brain are identified as being associated with various communication functions, it is important to note that these functions can be assumed by other brain regions. For example, a considerable degree of recovery can occur when Broca's area is injured since the surrounding regions share its specialization in latent form.

Human Acquisition of Language

Normal children are born with the ability and the drive to acquire the languages to which they are predominantly exposed from infancy. By late childhood the basic vocabulary of the "native" language has been acquired, together with its phonological and grammatical structure. The timetable of language acquisition derived from Lenneberg [Lenneberg 67] is shown in Table 6-1.

The behaviorist vs. innateness controversy. For a long time, scholars considered language acquisition to be carried out largely by analogy from observed patterns of sentences occurring in utterances heard and understood by the child. For example, Skinner [Skinner 57] incorporates the major aspects of linguistic behavior within a "behaviorist" framework, relating verbal behavior to variables such as *stimulus, reinforcement*, and *deprivation*, as used in animal experimentation. Subsequently, a number of lin-

TABLE 6-1 ■ Timetable of Language Development in the Child
3 months. When talked to, and nodded at, smiles, followed by voicelike gurgling sounds (cooing)
4 months. Responds to human sounds more definitely, turns head, eyes search for speaker, occasional chuckling sounds
5 months. Vowel-like cooing interspersed with consonantal sounds
6 months. Cooing changing into babbling, resembling one-syllable utterance
8 months. Distinct intonation patterns: utterances signal emphasis and emotions
10 months. Appears to wish to imitate sounds, but imitations are never quite successful
12 months. Definite signs of understanding some words and simple commands, sound sequence replicated
18 months. Definite repertoire of words, three to 50, much babbling, intricate intonation pattern, understanding progressing rapidly
24 months. Vocabulary of more than 50 items, two-word phrases, increased interest in language
30 months. Fastest increase in vocabulary, no babbling, utterances of at least 2 words, intelligibility not very good, but good understanding
3 years. 1000-word vocabulary: 80% of utterances are intelligible
4 years. Language well established

The timing of language development shown on this table is meant to be indicative; some children will progress faster and others slower than these milestones. (After E. H. Lenneberg. *Biological Foundations of Language*. Wiley, New York, 1967.)

guists led by Chomsky [Chomsky 75] have stressed the inherent grammar-building disposition and competence of the human brain, which is activated by exposure to language during childhood. In this point of view, no formal language instruction is necessary. One merely immerses the child in an environment in which the language is spoken, and the innate ability of the human brain to derive the appropriate grammatical structures and rules will automatically provide the child with linguistic competence.[11]

In examining the development of a child's language, one can find evidence supporting each of these views. The naming of objects with which the child is familiar and the association of *no* with disapproved behavior are examples of classical conditioning. Marshall [Marshall 80] discusses the body of evidence showing that the speech addressed to young children, termed "motherese," is typically different from that addressed to older children and adults:

> In the heyday of the "nativist" accounts of language acquisition [the early 1960s] it was widely assumed that the speech heard by children was a haphazard collection of sentence fragments, mistakes, backtrackings, throat clearings, and other kinds of unintelligible gibberish. There is now a considerable body of evidence showing that the speech addressed to young children is typically very different from that addressed to older children and adults.[p.115]

He cites recent work showing that some aspects of motherese are causally related to and can facilitate the rate of language acquisition.

Advocates of the Chomsky view point out that the *telegraphic speech* used by children is not a simple repetition of the adult's sentences. Thus, a parent will say "He is going out." but the child will convert this to "He go out." In general, parents do not seem to pay attention to such bad syntax; they do not even seem to be aware of it [Brown 77]. Study of complex sentences produced by children indicates that children apply their own grammatical rules (which are not direct imitations of adult grammar) in a systematic manner, and seem to acquire the conventional rules only through time and experience.

Recent experiments with infants young enough to be unaffected by their linguistic environment further support Chomsky's view that inborn knowledge and capacities underlie the use of language. Studies of speech perception in infants show that children have an innate perceptual mechanism adapted to the characteristics of human language [Eimas 75]. The research is based on the study of phonemes, the smallest units of speech that affect meaning. Phonemes are the auditory units that are analogous to the consonants and vowels of written language. We perceive speech "categorically," i.e., we are aware of the discrete phonemic categories, rather than of the continuous variation in the acoustic properties of sound.

Experiments with infants as young as one month can be carried out by measuring the rate that the infant sucks on a pacifier while being exposed to acoustic

---

[11]Note that this ability to learn by "immersion" disappears after approximately the first decade of life.

information. Increased sucking rate indicates the child's increased interest in a phenomenon. Another approach is to hold the child's attention with a toy, while a loudspeaker in another part of the room plays an acoustic signal. When a sound of interest occurs, the child turns in the direction of the loudspeaker.

It has been found that infants respond to phoneme categories rather than to the continuous gradations in acoustic properties of the signal. In one experiment, one group of infants was exposed to a phoneme sequence, another to an acoustic sequence, and a control group was exposed to no acoustic stimulus. In all of the groups, the sucking rate decreases at about the same rate. However after five minutes, a change in phoneme or acoustic signal is made. In the case of the group with the changed phonemic signal, the sucking rate increases to the original high value, while for the group with the changed acoustic signal the sucking rate continues to decrease.

Studies further show that all infants have the same inborn linguistic mechanism, but that the infant's linguistic environment causes the child to retain and improve perceptual capacities corresponding to phonemic distinctions in the parent language, while losing the ability to detect distinctions that do not occur in the native language. For example, English-background infants of six to eight months respond to Hindi consonantal contrasts, but lose this ability by age 10 to 12 months. Interestingly, the inactive perceptual mechanisms do not disappear completely, e.g., adult speakers of Japanese can, with enough experience, learn to distinguish the phonemes /r/ and /l/.[12]

"Carving up the world" into conceptual categories. People do not perceive the world as a continuum without any intrinsic boundaries. Rather, we partition the world into objects and categorize these objects as belonging to named classes. This classification allows us to relate new objects and events to classes of similar things with known properties.

Although such categorization seems to come quite naturally to us, we might be hard put to explain why we called the object in front of us a "bush." Do we have an image of a prototype bush stored in our mind to which we compared the object? On reflection, what might a prototype bush be? Do we have a list of bushlike properties, concerned with size of the object and shape of the branches and of the leaves? How do we asign the bush to the more general class or category of "vegetation?"

Though they cannot name the categories and relationships, very young children, 12 to 24 months old, have the ability to group and order objects on the basis of the various physical and functional relationships that hold among objects [Nelson 73]. Children of that age

[12]Categorical perception of speech sounds is not species-specific to the human. Other mammals, such as the chinchilla also have this ability, as do macaque monkeys [Flavell 85]. Aslin [Aslin 83] conjectures that since other mammals possess auditory categorical perception, this ability may have been acquired quite early in evolutionary history, before the capability for human oral speech. Thus, it is conceivable that human speech sounds are the way they are partly because our mammalian auditory system is constructed to readily discriminate and categorize these sounds.

group are first shown ten toy objects belonging to the same conceptual category, e.g., furniture. When they are shown a new pair of objects, a chair and an apple, they pay more attention to the apple than to the chair, since the chair is recognized as being a member of the category recently observed. The apple is attended to because of its novelty [Ross 80]. Even more remarkable is the fact that two-year-olds have been shown to possess scene schemas for how places look, e.g., what is to be expected in a kitchen scene [Mandler 83].

Flavell has written [Flavell 85]:

> Young children probably have representations of class-inclusion relations that are, in most important respects, not qualitatively different from those of older people. . . . However, they . . . are less able to talk and reason about class hierarchies and class-inclusion relations than older people are.

Subjects with damage to the posterior regions of the brain sometimes suffer from *nominal* or *anomic aphasia* in which they lose the ability to name and categorize objects. It has been suggested that this impairment is a result of disruption of associations involving different sensory modalities that are part of the naming act. Brown [Brown 80], relates the range of speech disorders ranging from phonological (production of speech) to semantic (meaning of the utterance) to the "triune" brain organization described in Chapter 2.

Animal Acquisition of Language

In the late 1960s, the Gardners of the University of Nevada published results indicating that a chimpanzee named Washoe was able to learn American Sign Language (ASL) [Gardner 69]. ASL was chosen to overcome the vocal limitations of the chimp. Washoe learned signs for hundreds of different objects and occasionally put together creative combinations of signs. (For example, the Gardners report that Washoe labeled a duck as a water bird.)

Herbert Terrace and his group at Columbia University attempted to duplicate this work by training a baby male chimp. After four years of work, they found that the chimp could indeed learn the American Sign Language (ASL) names of objects, but they claim that he could not reliably combine signs into grammatically correct sentences [Terrace 81]. They conclude that chimps cannot generally combine symbols to create new meanings. They also claim that analysis of videotapes made by their group, and by other groups, reveals that chimps often imitate signs made by humans, and this accounted for many of the "sentences."

Psychologist Francine Patterson at Stanford [Patterson 78] has reported that the achievements of Washoe have now been surpassed by Koko, a female gorilla trained since 1972.

A fierce controversy still rages between the animal language advocates and those who believe that any results indicating sophisticated language use or understanding by animals are due to unconscious clues given by the trainer.[13]

[13]This is called the "clever Hans" effect after a horse that was supposedly performing arithmetic computation, but was actually picking up very subtle cues from the trainer (apparently unbeknownst to the trainer).

Skeptics might be convinced of the possibility of animal language if the acquisition of a language enabled the animal to accomplish a nonlanguage-related task that it could not do before, e.g., if an animal with language skills could solve a problem while those without such skills would fail. To date, no one has attempted to demonstrate this. An anthology of important articles on both sides of the controversy is given in Seboek [Sebeok 80].

LANGUAGE AND THOUGHT

A study of American Indian languages led some scholars to speculate on the relationship between language, culture, and thought patterns. It was hypothesized that the world as mirrored in each language might have a strong effect on the perception and thought of the individual. Along these lines the linguist Edward Sapir has said: "We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation."

This idea was further developed by Benjamin Lee Whorf, and is now known as the Whorfian hypothesis [Whorf 56]. It is also known as the linguistic-relativity hypothesis because it proposes that thought is relative to the language in which it is conducted [Carroll 56].

> When Semitic, Chinese, Tibetan, or African languages are contrasted with our own, the divergence in analysis of the world becomes more apparent; and, when we bring in the native languages of the Americas, where speech communities for many millenniums have gone their way independently of each other and of the Old World, the fact that languages dissect nature in many different ways becomes apparent. The relativity of all conceptual systems, ours included, and their dependence upon language stand revealed.

Using the differences between Standard Average European (SAE) languages and the Hopi language, Whorf investigated the question, *Are our own concepts of 'time,' 'space,' and 'matter' given in substantially the same form by experience to all men, or are they in part conditioned by the structure of particular languages?* For example, the Hopi do not say "I stayed five days," but rather "I left on the fifth day," because the word day can have no plural. Whorf's conclusions [Carroll 56] are as follows:

> Concepts of "time"... are not given in substantially the same form but depend upon the nature of the language or languages through the use of which they have been developed.... Our own "time" differs markedly from Hopi "duration." ... Certain ideas born of our time-concept, such as absolute simultaneity, would be either very difficult to express or impossible and devoid of meaning under the Hopi conception.

Whorf found that there is a considerable difference between SAE concepts of 'matter' compared to Hopi, but that there was no great difference in the concept of 'space.'

In a later study [Carroll 56], researchers tested two groups of Navajo children, one group that spoke only English and the other only Navajo. In the Navajo language certain verbs of handling require special forms depending on the shape of the object being handled. The experiment compared the two groups with respect to how often they used shape, form, or material rather than color as a basis for sorting objects. It was found that the Navajo-speaking children tended to sort objects on the basis of form at significantly earlier ages than did the English-speaking children. The fact that the Navajo language required attention to shapes and forms seem to make the Navajo-speaking children pay more attention to this aspect of their environment.

The Whorfian hypothesis is still a subject of debate: Alford [Alford 78] has surveyed criticisms of the hypothesis, and Malotki [Malotki 83] has recently carried out a deep analysis of Hopi that disagrees with Whorf's conclusions concerning the temporal concepts.

A widely held view is that there is indeed a correspondence between language and the ways of conceiving the world, but that language differences are caused by the experiences or needs of a particular people, rather than by the dictates of some arbitrarily defined linguistic system—i.e., it is ultimately the physical environment rather than the arbitrary choice of language that structures our thought processes. Thus, because it is important for Eskimos to be able to describe the different types of snow and ice, they will create different words for these.

As indicated by Rosch [Rosch 77], our categorization of the world is not arbitrary; it depends on information in the natural world to which we as a species are geared to respond (see Box 6-1). In this view, language, for the most part, follows upon discriminations made by individuals rather than playing a controlling role in how one classifies the world.

BOX 6-1 Natural Categories and Natural Kinds

The Sapir-Whorf hypothesis and related investigations imply that the language we use critically affects our view of the world and how we are able to think about things in the world. It is therefore important to determine the extent to which differences between languages are arbitrary, and the extent to which similarities are accidental. In particular, how and why does a language "carve up the world" in a particular way?

There had been a long standing belief that (1) the common objects of the world can be classified into distinct groups; (2) these groups can be defined by specific criterial attributes which are relatively independent; and (3) people speaking different languages made different distinctions, i.e., formed different categorizations suited to their particular needs. Thus, it was noted that while the color spectrum is continuous, every human culture has a somewhat different way of breaking it up into named color categories; some cultures employ only two or three named colors. A person growing up within a culture learns the color names that have been arbitrarily chosen by that culture.

In the 1970s Eleanor Rosch performed experiments which demonstrated that members of different cultures remember and make color similarity judgments that are indistinguishable from each other in spite of significant language differences (see Chapter 12 of Gardner [Gardner 85]). Based on these experiments and other observations Rosch concluded that:

1. Naming practices of cultures are relatively unimportant compared to the innate organization of the human nervous system in making category judgments.

COMMUNICATION

The Mechanics of Communication

To communicate with a person or machine, the receiver must know that you are communicating with it, must be willing to listen, be able to understand your language, and possess a similar social and conceptual "frame of reference." To be most effective, the receiver must have some way of signaling success in understanding your message, and you must be able to tell after a while whether or not the receiver understands what you are trying to say. Most of these requirements are fulfilled automatically when we speak to another person. We attract the other person's attention, we note by physical cues whether or not the person is listening, and the person signals understanding by nodding or by responding with a communication. We can tell whether the listener understands by analyzing the responses. When people realize their messages are not being properly understood, they modify or terminate their conversation.

There are also cultural assumptions. Once we realize that a person speaks our language and shares our cultural framework, we can make a point using an expression such as *A rose by any other name smells as sweet,* and assume that the person understands that we are trying to indicate that the intrinsic properties of an object are not altered by the name we assign to it. If not given the requisite knowledge and reasoning ability, a machine might treat this as an ordinary sentence and place in its database some statement such as *The odor of a rose*

BOX 6-1 (*continued*)

2. There is considerable redundancy in the appearance of members of natural categories (e.g. birds, trees, flowers)—their defining features are not independent; our recognition mechanisms exploit these redundancies.
3. Human categorization is more closely linked to similarity to an exemplar or prototype of a class than to the presence of a fixed set of features.
4. Categories in the real world are not sharply defined, but blend into one another.

Other attacks on the classical views of concept and category came from Wittgenstein [Wittgenstein 68] who felt that concepts are neither mental constructs in the head nor abstract ideas in the world, but rather are community-developed tools for accomplishing things. Putnam [Putnam 75] believed that the world is not a perceptual jumble that can be arbitrarily partitioned, but rather that there are inherent structures, "natural kinds," that allow us to form stable generalizations and then reason about things in the world.

Objections to the classical view of language and category can be summarized as follows: To deal effectively with their environment, people form linguistic categories for things that appear similar or behave in similar ways; such categorizations reflect the perceptual structure of the perceiver and are not arbitrary. Nevetheless, people also form categories far removed from direct perceptual observation, e.g., the categories of odd and even numbers. Such categorizations, essential to human cognition, more closely follow the classical view.

is not affected by the name assigned to it.

People use various methods of communicating commands, questions, anger, information, promises, threats, belief, and desires to other people. They communicate not only by written and spoken language, but also by *body language,* nonverbal communication involving body posture, facial expression, seated position, and other body signs. Such nonverbal communication, though very subtle, can be interpreted with great accuracy.

Often the same words can mean different things depending on the social setting, or the tone and intonation used. For example, the question, *Are you going to get the book?* can be used as a question, as a threat, or as a command.

Vocabulary of Communication

The vocabulary used by people is much smaller than one might expect. The following tables are for English, but they are about the same for French, Russian, and many other "natural languages,"especially those employing phonetic alphabets. Table 6-2 shows that only a small portion of the words in an abridged dictionary are commonly known; an average adult's vocabulary consists of about 10% of all dictionary words.

TABLE 6-2 ■ Size of Vocabulary Employed by Various Sources

Source	No. of Words
Child	3,600
14-year-old	9,000
Adult	12,000–14,000
Abridged dictionary	150,000
Dante's *Divine Comedy*	5,900
Homer's poems	9,000
Shakespeare's works	15,000–25,000

After A. Kondratov. *Sounds and Signs.* MIR Publishers, Moscow, 1969.

From Table 6-3 we see that with a 3000-word vocabulary we can expect to recognize 90% of the words on a page of general text. A 1000-word spoken vocabulary will allow the same recognition of spoken words. Computer understanding of language would be simple if language understanding was merely a matter of looking up word meanings, since the required vocabulary is not large in terms of computer memory. The next section will point out the reasons why understanding is far more than the stringing together of individual word meanings.

TABLE 6-3 ■ Frequency of Use of Spoken and Written Language Words

Spoken Language Vocabulary	Written Language Vocabulary	Probability of Appearance of Words in Speech or Text
750	—	75.0%
—	1000	80.5%
—	2000	86.0%
1000	3000	90.0%
2000	5000	93.5%

After A. Kondratov. *Sounds and Signs.* MIR Publishers, Moscow, 1969.

Understanding Language

As indicated above, language understanding is a form of reasoning in which the intended communication of the source is deduced from the combination of the spoken or written message, the recipient's intuitions as to the "state of mind" of the source, the context, knowledge of the language, and knowledge about the world. When one examines the problems involved in understanding a natural language expression, one wonders how people are able to learn language as children, and how the thought encoding and understanding process comes so effortlessly. For example, consider the following two sentences:

1. I have a headache tonight.
2. I will have a headache tonight.

The surface meaning of the first sentence is that the speaker is feeling ill, but the deeper meaning can be a refusal to be sociable or romantic. The second sentence, because prediction of illness is usually not possible, would be considered an insulting refusal.

Many of the sentences that are used by people are ambiguous in some way, but people are so facile at decoding the meanings that the ambiguities often go unnoticed. For example, the sentence *Time flies.* would not be considered to be ambiguous since most people would see only the statement *Time passes quickly.* and not the command *Determine the flight speed of a set of insects!* In addition, people are unaware of how much general knowledge is often required to understand even simple sentences in natural language. In the following sentences supplying the word or idea referred to by *It* requires knowledge about a variety of objects in the world:

The car ran over the toy in the driveway.

It shouldn't have been there.

It was scratched and had to go to the garage.

It was scratched and had to be repaved.

It was too bad.

Notice that each of the *Its* refers to a different aspect of the original sentence. The first *It* refers to the toy, because we know that cars belong in driveways whereas toys do not. The next *It* refers to the car, because we know that scratched cars can be fixed in a garage. The next *It* refers to the driveway, because driveways are repaved. Finally, the last *It* refers to the whole incident. A remarkably broad knowledge database is required to supply the proper referent for each of the *It* terms.

An even more sophisticated level of reasoning is needed to understand sentences such as: "Mary wondered why everyone was driving under 55, and then she saw."... "The man handed the teller a note and she pressed the silent alarm." The first sentence requires the knowledge that the number refers to 55 miles per hour, that the speed limit is 55 miles per hour, that if one exceeds this limit one can get a speeding ticket that costs time and money, and that therefore drivers heed this limit when a police car is close by. The second sentence requires the knowledge that the teller is a bank teller, that bank robbers often hand a note demanding money to the teller, and that help can be summoned by pressing the silent alarm.

Spoken Language. Spoken language has an additional problem that must be overcome to understand an utterance. When people speak, they run their words together so that, for example, someone who does not speak English might hear "Didja sayuwer goin?" instead of "Did you say you were going?" Thus, a person who does not speak English would not be able to understand what was said by listening to that sentence, writing down the words, and then looking them up in a dictionary. "Didja" and "sayuwer" are not in the dictionary. The separating of words, called "segmentation," requires an understanding of the language. Therefore, a computer designed to understand spoken language must be provided with rules that indicate how to segment the words (plus all of the other knowledge that it takes to interpret the utterance).

Sign Language as Language. Language extends beyond the obvious spoken and written forms. We immediately think of sign languages such as American Sign Language ("Ameslan" or ASL), British Sign Language, etc. that are, in the words of Oliver Sacks [Sacks 86],

> . . . . complete in a 'Chomskian' way. Their syntax and grammar are complete, but have a different character from that of any spoken language. Thus it is not possible to transliterate a spoken tongue into Sign, word by word or phrase by phrase—their structures are essentially different. It is often imagined, vaguely, that sign language is simply [an alternate version of the user's native tongue such as] English or French: it is nothing of the sort; it is itself, Sign. . . . Sign language enables its users to discuss any topic, concrete or abstract, as economically and effectively as speech.[14]

Speaking is an ability that must be taught to the deaf, and it takes many years. On the other hand, the deaf show an immediate and powerful disposition to sign. Deaf children whose deaf parents use sign language make their first signs when they are about six months old and have considerable sign fluency by the age of 15 months. As Sacks says, "This is intriguingly earlier than the 'normal' acquisition of speech, suggesting that our linguistic development is, so to speak, retarded by speech, by the complexity of neuromuscular control required. If we are to communicate with babies, we may find that the way to do so is by Sign."

A child can become fluent in signing by the age of three years, and then can acquire reading and writing, and even speech. There is no evidence that signing inhibits the acquisition of speech, and the reverse is probably true.

David Wright [Wright 69], who became deaf at the age of seven years, provides an interesting insight into the role of spoken language in the development of childrens' world knowledge. He comments on a congentially deaf schoolmate, "She was far from stupid; but having been born deaf her slowly and painfully acquired vocabulary was still too small to allow her to read for amusement or pleasure. As a consequence there were almost no means by which she could pick up the fund of miscellaneous and temporarily useless information other children unconsciously acquire from conversation or random

[14]Prior to 1750 there was no hope of literacy or education for most of those born deaf. The Abbé de l'Épée founded the first school for the deaf in 1755.

reading. Almost everything she knew she had been taught or made to learn."

MACHINE UNDERSTANDING OF LANGUAGE

The goal of computational linguistics is to develop theories and techniques that would allow a computer to derive *meaning* from natural language expressions, and produce written or spoken natural language. Determining the intent of a message, rather than only its literal content, requires a combination of language-specific and general common-sense reasoning mechanisms. In the most sophisticated applications, a language understanding program must model the beliefs and knowledge of the agents participating in the communication and be able to deal with incomplete and sometimes inconsistent information.

Faking Understanding

If the responses from a computer seem reasonable, people will tend to ascribe more understanding to the computer than actually exists. A classic example of this is Weizenbaum's ELIZA program [Weizenbaum 66], which takes the role of a nondirective psychoanalytic therapist. The role is relatively easy to imitate because this type of psychoanalysis elicits the patient's responses by reflecting his statements back to him. ELIZA uses a set of stored word patterns, such as:

INPUT: I am _________.
OUTPUT: Why are you _________?

ELIZA merely uses the words that appear in the _________ portion of the input to construct the output. Thus, an input of *I am very sad* will result in an output of *Why are you very sad?*. If the words "mother,""father,"" brother," or "sister" appear in the input, an ELIZA canned response might be *Tell me more about your family*. When ELIZA cannot find a word pattern that matches, it responds with ambiguous phrases designed to elicit further responses from the user, such as: *Tell me more. In what way? Can you think of a specific example?* It is simple to include features so that "canned" phrases are not repeated during a session, and to provide phrases in random order so that no fixed pattern of response is detected by the user.

A typical ELIZA dialogue is:

All men are alike.
IN WHAT WAY?
They are always bugging us about something or another.
CAN YOU THINK OF SPECIFIC EXAMPLES?
Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE?

ELIZA was so effective that people used the program to seek advice, even when they were told of its internal structure and that it lacked *any* mechanism for understanding. This ready acceptance of ELIZA as a real therapist motivated Weizenbaum to warn of the dangers of applying the computer to areas requiring human judgment [Weizenbaum 76].

What Does it Mean for a Computer to Understand?

It is very clear that, given its simple template-matching design, the ELIZA program does not in any sense understand its input or output. However, as we devise more sophisticated computer programs for dealing with natural language, we are faced with a deep philosophical question: What kind of ability to manipulate the written or spoken symbols of a language amounts to a *true understanding* of that language? A thought experiment, "the Chinese room," by the philosopher John Searle [Searle 84] vividly captures the problem of computer understanding. Searle states that he understands no Chinese at all and can't even distinguish Chinese symbols from some other kinds of symbols. He imagines that he is locked in a room with a number of cardboard boxes full of Chinese symbols, and is given a book of rules in English that instruct him how to match these Chinese symbols with each other. The rules tell him that a certain sign is to be followed by a certain other sign. The people outside the room pass in more Chinese symbols and, following the instructions in the book, he passes Chinese symbols back to them. Unknown to him the people who pass him the symbols call them *questions* and the book of instructions that he works from they call *the program*; the symbols he gives back to them they call *the answers to the questions* and they call him *the computer*.

> Suppose that after a while the programmers get so good at writing the programs and I get so good at manipulating the symbols that my answers are indistinguishable from those of native Chinese speakers. I can pass the Turing test for understanding Chinese. But all the same I still don't understand a word of Chinese and neither does any other digital computer because all the computer has is what I have: a formal program that attaches no meaning, interpretation, or content to any of the symbols. . . . What this simple argument shows is that no formal program by itself is sufficient for understanding, because it would always be possible in principle for an agent to go through the steps in the program and still not have the relevant understanding.

Hofstadter [Hofstadter 83] answers as follows:

> Our response to this is basically the 'systems response,' that it is a mistake to try to impute understanding to the (incidently) animate simulator; rather it belongs to the system as a whole, which includes what Searle characterizes as a 'few slips of paper.'

To the "system" advocates, Searle suggests that the person in the room should simply memorize or incorporate all the material in the few slips of paper. The systems people retort that a key part of Searle's argument is in glossing over questions of order of magnitude and that nearly all of the understanding must lie in the billions of symbols on paper.

Others say that "understanding" is achieved by an entity when (a) it has adequately modeled some situation of interest, and (b) this restricted model is strongly linked to the "world model" of the entity, i.e., many, or most of the relevant associations have been explicitly established. In the Chinese room example, condition (b) has not been satisfied, and indeed, to the extent that a "computer entity" consists of a disconnected set of models, performance does not imply understanding. However, if the computer has a sufficiently rich integrated world model, then any reasonable *operational definition* of the term "to understand" will be satisfied. Note that this view disagrees with Hofstadter—the critical factor is not the complexity of some restricted model of interest, but rather the connections of the restricted model to a comprehensive world model.

A related but distinct point of view (see Box 6-2) holds that words and sentences are not ultimately definable in terms of an objective world, but that every reading or hearing of a text constitutes an act of giving meaning to it through interpretation. Interpretation depends on a person's tradition or preunderstanding and as people experience the world their understanding changes as does the meaning they derive. Thus meaning is not a linkage between text and reality, but rather a dynamic coupling between users of a common language.

And so the discussion rages back and forth. The question of what constitutes "real understanding" will become increasingly pertinent as machines become more competent and assume a greater decision making role in human affairs.

The Study of Language

Language can be examined from many different points of view, including the study of language universals, language acquisition and use, and philosophy of language, to name only a few. We will be concerned here with the following aspects of language because of their relevance to computer understanding: (1) syntax, the study of sentence structure; (2) semantics, the study of meaning; and (3) pragmatics, the study of the uses to which language is put and how speaker's goals are achieved by uttering sentences in context. While this partition is useful for discussion purposes, it should be kept in mind that there is not always a clear line separating these topic areas. Winograd [Winograd 74] uses the analogy of a jigsaw puzzle to explain the role of syntax, semantics, and pragmatics.

> The shape of the jigsaw pieces might correspond to the syntax of language—

BOX 6-2 A Philosophy of Understanding

Research in understanding of language assumes that meaning is derived from a string of words and their context. This assumption of a unique meaning is challenged by the philosopher Hans-Georg Gedamer [Gedamer 76] who insists that interpretation depends on a person's tradition or pre-understanding; as people experience the world their understanding changes and the meaning they derive changes. Winograd [Winograd 86] indicates the relevance to AI of hermeneutics, the science and methodology of interpretation,

> In a way, frame-based computational systems approach meaning from a hermeneutic direction. They concentrate not on the question, 'How does the program come to accurately reflect the situation?' but rather 'How does the system's pre-knowledge (its collection of frames) affect its interpretation of the situation?' The meaning of a sentence or a scene lies in the interaction between its structure and the preexisting structures of the machine.

Unlike a person who can modify existing mental frames or incorporate new ones based on experience, at the present time the computer program must rely on its designer for these modifications. Without this ability to change, the machine will not be able to make increasingly mature interpretations.

UNDERSTANDING LANGUAGE →

message — SYNTAX — SEMANTICS — PRAGMATICS — intended meaning

← GENERATING LANGUAGE

FIGURE 6-3
The Approach to Computer Understanding and Generation of Language.

> there are rules for how the different shapes fit together and some pieces can be assembled without regard to what appears on them. . . . We might view things like color and texture as a kind of simple picture semantics which indicates what sorts of elements can fit with others. . . . Finally, there is a more sophisticated pragmatics or reasoning based on knowledge of pictures. If a picture of an elephant is emerging, it might be useful to look for something with the color and texture of an elephant tail, and then use its further color and shape information to guide the process.[p. 46]

The role of syntax, semantics, and pragmatics in understanding and generating language is shown in Fig. 6-3. In the understanding-language direction, the structure of the message, derived by syntactic analysis, is processed semantically to extract the literal meaning of the sentence. A pragmatics analysis derives the "intended meaning" by using world knowledge, knowledge of the context, and a model of the sender. The process operates in reverse in language generation. We show the connections as dashed lines to avoid giving the impression that the process is necessarily a sequential one. There are some approaches that blur the distinction between syntactic and semantic analysis, and some that deal with syntax, semantics, and pragmatics in parallel, moving back and forth from one to the other as the analysis proceeds.

Natural language offers a remarkable palette that enables people to communicate information about objects, actions, beliefs, intents, and desires that occur over time and space. The nuances of meaning must be captured by the computer if it is to have the linguistic power possessed by people. However, for a computer to deal with the "meaning" of natural language expressions, it must convert the things portrayed by natural language to a form that is amenable to computer manipulation under the guidance of a formal set of rules. Ideally, the transformation from natural language to a computer representation would provide a means for the computer to reason about the information, deal with questions, take requested actions, and make appropriate responses. The present state of the art is far from "ideal" since we do not know how to represent in a computer the full scope of meaning that is indicated above. Instead, the available representations are limited to the modeling of meaning in specialized domains such as storage and retrieval of information in an application-specific database, interacting with people in constrained situations, and answering questions about, or paraphrasing, a given segment of focused text.

Syntax. No speech community has ever been identified where communication is restricted to single-word discourse. Instead, words are concatenated (strung together), and we know of no language where words are strung together randomly. It is generally assumed that there must be a finite set of rules that defines all grammatical operations for any given language. Any native speaker will generate sentences that conform to these grammatical rules, and any speaker of the speech community will recognize such sentences as grammatical.

The study of syntax is a fascinating one in its own right, having widespread implications that range from ideas on language universals to conjectures about language acquisition by children. For our purposes, we will take a more limited point of view and consider that the purpose of syntax is to provide a structural representation that will be useful in the understanding process.

Computational linguistics attempts to develop formal rules that assign structural descriptions to sentences in an explicit and well-defined manner. To indicate the nature of a formal approach, consider a simple *phrase structure grammar* using rules of the type,

Sentence → NP + VP,

where → stands for "is made up of," + stands for "followed by," and the capitalized letters stand for category symbols such as "sentence," "noun phrase," or "verb phrase." Thus, the rule says *a sentence is made up of a noun phrase (NP) followed by a verb phrase (VP).*

The primitive set of "rewrite rules" given below is indicative of the machinery of a phrase structure grammar:

1. Sentence→NP + VP
2. NP→T + N
3. VP→Verb + NP
4. T→the
5. N→(man, ball . . .)
6. Verb→(hit, took, . . .)

Note that a rule such as rule 2 can be interpreted either descriptively (declaratively), *A sentence is a noun phrase followed by a verb phrase,* or as a procedure that says *If you want to find a noun phrase, look for "the" followed by a noun.*

The rules can be used to analyze the phrase structure of a sentence, as shown in Fig. 6-4.

The procedure of Fig. 6-4 is called *bottom-up parsing.* The part of speech of every word is found in a lexicon or dictionary, and then the rules are used

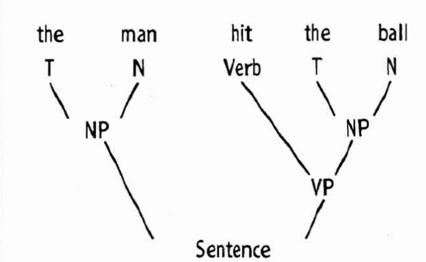


FIGURE 6-4 A Parse Tree.

to join the parts of speech together into phrases. In the bottom-up approach, much time could be spent examining combinations that are not legal. An alternative approach, known as *top-down parsing*, attempts to find instances of given rules in the text. Thus, one would look for verb followed by noun phrase because we know that this forms a verb phrase. In this approach, one searches for what is expected or wanted in the way of structure. Sophisticated parsers use a combination of top-down and bottom-up strategies.

A major problem in parsing is caused by the fact that many words have multiple part-of-speech assignments. The multiple assignments lead to many possible structures since, as the analysis of the sentence proceeds from left to right, many potential phrases must be retained until further words are encountered that show that a particular phrase structure is or is not possible. For example, in parsing the sentence "The table covers were soiled," it is necessary to consider "covers" as both a noun and a verb, so that after scanning the leftmost three words we would have both (the table)(covers) and (the table covers) as structural partitions. Only when the word "were" is encountered can we drop the first parsing.

The structural analysis of a typical sentence is far more complex than our example might imply, as shown in the parse tree of Fig. 6-5. Note that many more word and phrase classes are used in comparison to the simple example presented in Fig. 6-4. This additional structure is necessary to represent the more complex relationships among words and phrases found in most natural language expressions. However, because of its limited expressive power, the rewrite-rule formulation is not suited to describing procedures needed for sophisticated parsing. For example, it is difficult to express the constraints that must be satisfied among various parts of speech, such as agreement between subject and verb. Further, rewrite rules are unable to express high-level guidance as to the strategies to be used in parsing a sentence. Therefore, other more general representations have been developed; Appendix 6-1 describes two such representations.

In addition to trying to represent the complexities of the parsing process, it is important to be able to describe efficient parsing techniques. The approaches shown in Appendix 6-1 are known as "nondeterministic parsers" because the parser makes a best guess at any particular stage, but may have to back up if the guess turns out to be wrong. The chart parser representation described in Appendix 6-1 is especially suited for describing efficient backtracking procedures. Another approach is that of Marcus [Marcus 80, Thompson 84] who believes that much of the effort expended by ordinary parsers is due to the multiple parsings resulting from local structural ambiguity, rather than the ambiguity of the full sentence. He feels that people use a single local parsing that follows from the partial structure and the next four or five words, rather than developing all possible local parsings and then choosing the most applicable one. Using this approach, his "deterministic parser" stores fragments of the syntax tree in several temporary buffers, and uses a set of rules to determine the most likely local parse.

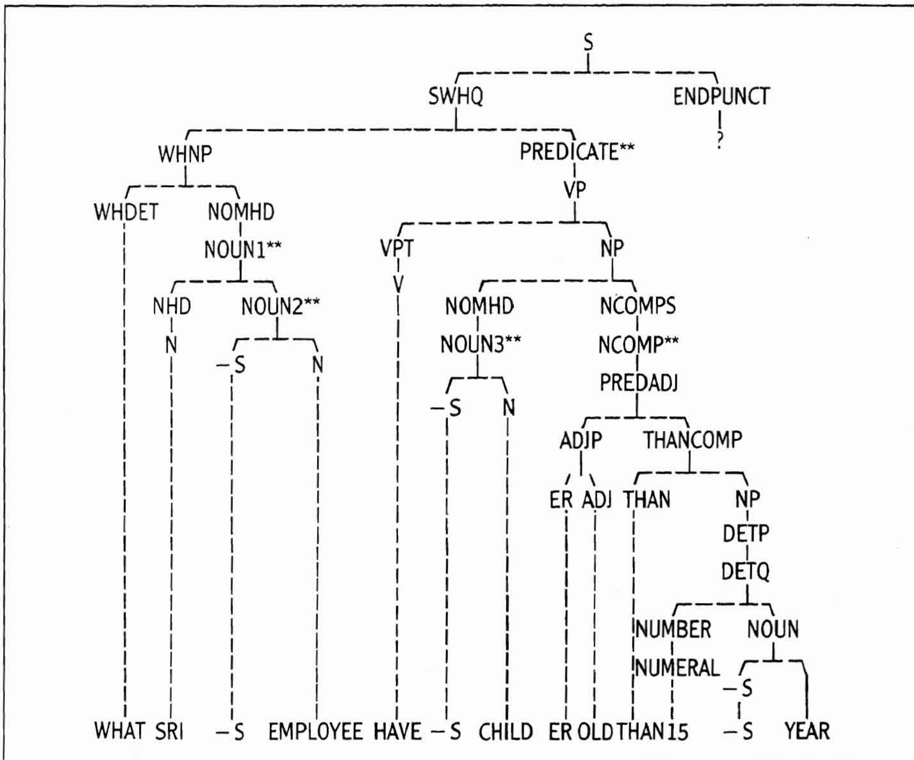


FIGURE 6-5 Output from DIALOGIC, a Sophisticated Parser.

Parse tree for "What SRI employees have children older than 15 years?" WHNP = interrogative noun phrase; WHODET = interrogative determiner; NHD = pronoun modifier head; SWHQ = what question; VP = verb phrase; NP = noun phrase; NCOMPS = noun-phrase complement; PREDADJ = predicate adjective; DETQ = determiner/quantifier phrase.

Semantics. The question of meaning is a deep philosophical one. For our purposes here we will consider the derivation of meaning as the problem of converting a sentence to a representation that can be related to real or imaginary objects in a possible world. Semantics will be used as the basis for expressing the literal meaning of an expression; the intended meaning will be derived using pragmatics.

There are two main approaches to the assignment of literal meaning to an expression. The first is "lexical" semantics that gives prime importance to *content* words. By assuming that such words have a direct relationship to "deeper" notions, the lexical semanticist hopes to show how words fit together. The representation of the semantics of a sentence is a paraphrase in which the content words of the original sentence have been rewritten and fitted together using their generic representations. The paraphrase is the program's "understanding" of the input.

In the "compositional" semantics approach, the meaning of a complex expression depends on the meaning of its subexpressions. Thus, the analysis of a phrase is its translation into formulas of an appropriate logical calculus. This is accomplished by using rules that describe how to bring together the formulas of subphrases of the phrase while taking into account the context of the phrase. Compositional semantics attempts to provide logical descriptions of how a phrase or word modifies another. The phrase, translated into a logical expression, is usable in a formal deduction system and forms the basis for any next step in the meaning derivation process, perhaps involving the pragmatics of the situation.

Lexical Semantics. Conceptual dependency (CD) theory is an example of a nonformal lexical semantics approach based on a set of elementary concepts (semantic primitives) that are used to express the meaning of an utterance. Schank and his colleagues [Schank 81] maintain that an extremely small set of primitive actions will account for what must be represented in the physical world. The CD representations should be identical for different sentences that describe the same event in quite different linguistic constructions. Some primitive actions such as *transfer, propel,* and *ingest* are shown in Box 6-3, and cover many simple physical events and human interactions.

In the CD approach, meaning is considered to be the primary issue, and the study of syntax is guided by the de-

BOX 6-3 The Conceptual Dependency (CD) Approach

The conceptual dependency approach uses a very small number of primitives to represent the actions of the physical world. Some of the key conceptual dependency primitives are presented below, using simple sentences as examples:

John went to New York.

actor: john
action: PTRANS (physical transfer of location)
object: John
direction TO: New York
FROM: unknown

The rock hit the boy.

actor: unknown
action: PROPEL
→ CAUSED { state: contact; object1: rock; object2: boy }
object: rock
direction: TO: boy
FROM: unknown

John drank a glass of milk.

actor: John
action: INGEST
object: milk
direction: TO: mouth of John
FROM: glass

instrument used (level I)

actor: John
action: PTRANS (physical tranfer)
object: glass containing milk
direction: TO: mouth of John
FROM: table

instrument used (level II)

actor: John
action: MOVE (movement of part of body)
object: hand of John
direction: TO: glass
FROM: unknown

instrument used (level III)

actor: John
action: GRASP
object: glass of milk
direction: TO: hand of John

mands of the theory of understanding. Originally there was not to be an independent syntactic pass. Over the years this restriction has been relaxed, but semantics still tends to be the main focus. Every "content word" is considered to have an associated set of "slots," variables whose values are to be established, such as *actor*, an *action* performed by that actor, an *object* that the action is performed on, and a *direction* in which the action is oriented. This form forces the system to postulate actors, objects, etc. that may have not been explicitly mentioned in a sentence describing an event, but which nevertheless must exist. When converting a sentence into CD form, one of the prime motivators of the system is *slot-filling*, i.e., instantiating the actor, object, . . . slots by the words of the text.

The CD approach would translate the sentence "John went to New York" to a representation that indicated that the content word "went" is a primitive action known as PTRANS (for physical transfer of location). The "actor" John carried out PTRANS, the "object transferred" was John, the "direction-to" was New York and the "direction-from" is unknown. The word 'went' motivates the search in the sentence for the actor, object, and direction to fill the "slots" in the PTRANS semantic primitive structure.

Scripts were developed by Schank's group as additional methods for representing the context of a discourse. The script, a description of an event such as going to a restaurant, serves to organize the knowledge that people must have to understand some coherent segment of

BOX 6-3 (*continued*)

Here we see the potentially unbounded expansion of instruments. The analysis reads "John drank the milk by getting the milk to his mouth by moving his hand to the milk and grasping the glass."

The CD work has been applied to SAM [Schank 81], a program that answers questions about stories. Given the input "John went to a restaurant. He sat down. He got mad. He left," the program produces:

> John was hungry. He decided to go to a restaurant. He went to one. He sat down in a chair. A waiter did not go to the table. John became upset. He decided he was going to leave the restaurant. He left it.

Note that John could have gotten mad for various reasons, but the script used to help interpret the above text only provided one reason for getting mad, namely that the waiter did not come.

Another CD work is PAM [Schank 81], a program that reasons about people's intentions. Given the story,

> John loved Mary but she didn't want to marry him. One day, a dragon stole Mary from the castle. John got on top of his horse and killed the dragon. Mary agreed to marry him. They lived happily ever after,

PAM produces,

> John was in love with Mary. She did not want to marry him. A dragon took her from the castle. He learned that that dragon had taken her from the castle. He mounted a horse. It took him to her. He killed the dragon. She was indebted to him. She told him that she was going to marry him. He married her. He and she were happy thereafter.

Other CD efforts described by Schank and Riesback [Schank 81] are TALE-SPIN, a program that writes simple stories, and POLITICS, a program that simulates human ideological understanding of international political events. In all of these, natural language sentences are converted to CD form, and the various scripts and plans are used to direct the slot-filling operation.

human experience. In addition, scripts point out what behavior is appropriate for a particular situation. For example, a subway script would specify the participants (riders, cashier, conductor), the objects (turnstile, train, seat), and the episodes (getting a subway token, going through the turnstile). Understanding a story first requires the determination of what script is referred to in the story, i.e., setting up a correspondence between the vocabulary of the script and the story. (Note that this brings us face to face again with the problem of relevance, "how does a system know which script to choose?") Next, that script is used to identify and fill in the important details in the causal chain being built.

Compositional Semantics. The term "compositional semantics" is used to indicate a system in which the meaning of a complex expression depends only on its subexpressions. This permits every well-formed subexpression to be used as the basis for meaning-dependent processing. The computational linguistics approach to compositional semantics has been to derive a logical form from natural language expressions, and to use this form as the basis for deriving the meaning of the expressions. When a sentence is ambiguous the analysis must furnish distinct logical form representations for the different readings. The logical form is used as the intermediate step between the original sentence and the final expression that captures the meaning.

The essential problem for a theory of logical form is to represent in a logical formalism specific concepts of natural language such as events, actions, and processes; time and space; collective entities and substances; and propositional attitudes ("believe") and modalities ("should"). The theory is concerned with the question of what particular predicates, functions, operators, and the like are needed to represent the content of natural language expressions. Moore [Moore 81] surveys key problems that arise in representing the content of English expressions. Two typical examples of the type of problem that Moore discusses are:

- How can one reconcile statements that refer to points in time with those that refer to intervals? ("The company earned $5 million in March" does not mean that at every point in time during March the company earned $5 million.)
- How can one deal with collections? ("Newell and Simon wrote *Human Problem Solving*" does not mean that they each did it simultaneously.)

DIALOGIC [Grosz 82], is an example of a system that translates English sentences into logical form representations. Given the question, "What SRI employees have children older than 15 years?," the parser first produces the tree shown in Fig. 6-5. The system then converts the tree to a logic formalism after assigning additional attributes to nodes in the tree, identifying the quantifiers, heads of noun phrases, verb phrases, and adjectives. The final logical form is assembled by a procedure that determines the scope of the quantifiers and takes into account the characteristics of the database to be searched.

The final expression, shown in Fig. 6-6, can be paraphrased as, "Who is each employee such that the company of the

LOGICAL FORM	EXPLANATION
[QUERY	;Sentence is a query
(WH employee1	;Find someone in the ;database, say employee1
(AND	
(EMPLOYEE employee1)	;who satisfies the ;predicate EMPLOYEE
(EMPLOYEES-COMPANY-OF employee1 SRI))	;AND is employed by SRI
(SOME child2 (CHILD child2)	;There exists someone ;say child2 who satisfies ;the predicate CHILD
(AND	;AND also satisfies the ;predicate CHILD-OF for
(CHILD-OF employee1 child2) ((*MORE* OLD) child2 (YEAR 15))]	;the variable employee1 ;and child2 satisfies ;MORE OLD than 15

FIGURE 6-6 Logical Form Obtained by the DIALOGIC System.

employee is SRI and some child of the employee is older than 15 years?" This transformation of the original sentence is far from trivial, since the predicates relevant to the database such as EMPLOYEE, EMPLOYEE-COMPANY-OF, and CHILD must be identified with the words of the question, and the quantifier SOME and its scope has to be determined.

To obtain an answer to a question, the logical form of the question is considered to be a theorem to be proved, using the database (expressed in formal logic, also) as the set of axioms. Other natural language programs that interact with a retrieval database are described briefly in Box 6-4.

Comparing Lexical and Compositional Semantics. As described in the discussion on representation (Chapter 3), there are three main components of a representational system: (1) the "vocabulary" of the representation,[15] (2) the models based on this vocabulary which describe the structures and relàtionships among the things in the world and can be used to predict behavior, and (3) the symbolic formalism and the physical encoding that is used in the computer.

The lexical semantics approach uses a vocabulary of about ten to twenty basic concepts, some of which are complex enough to be considered models. Additional models are provided to capture more complex activities in the world, such as going to a restaurant, taking a train,

[15]In the case of a formal representation for natural language, this vocabulary could include both natural language and logic terms.

BOX 6-4 Question-answering Systems

Much work has been carried out in relating queries, written in a limited subset of natural language, to a database of facts about some limited domain. To answer questions, the system uses the statements contained in the database to reach a conclusion that fits the question. This box provides brief descriptions of two natural language "front ends" for retrieval systems. Both convert a query to a logical expression for use in searching the database.

The LUNAR System

The LUNAR system [Woods 77] is an example of a question-answering system in which the parser provides a structural description of the question, and interpretation rules identify the logical connections among the linguistic elements that correspond to database entries. Retrieval operations are performed using the query expressions produced by the interpretation rules. Thus, LUNAR transforms the question, "Do any samples have greater than 13 percent aluminum oxide?" into the expression,

```
(TEST (FOR SOME X1 ((SEQ
        SAMPLES) CONTAIN X1
        (AL203)
        (GREATERTHAN 13 PCT))))
```

The LUNAR database contained chemical data on lunar rock and soil composition from the Apollo moon missions. A question such as "What is the average concentration of aluminum in high alkali rocks?" would first be parsed. The phrase "high alkali rocks" would be found to correspond to a set of entries in the database, and 'aluminum' would be identified as one of the attributes. The phrase "average concentration" would be recognized as a particular set of computations that the system knew about, the computation would be made, and the answer given to the

what to do if one needs money, etc. In compositional semantics, the vocabulary used is that of formal logic, plus certain of the words in the original sentence. To make the logical form refer to something in the actual world we must supply additional assertions about the actual world, i.e., we must supply a model of the world of interest. Thus, the same logical form could mean different things, depending on the world model being used.

Each of the approaches to semantics has its strengths and weaknesses. The lexical approach is not general, and therefore tends to have an unlimited growth of special situations. However, its use of frames as a focusing device is very effective computationally, and it is robust with respect to ungrammatical sentences. Some of the major problems in script-based parsing include (1) indexing difficulties with a large database of scripts, (2) the problem of having multiple scripts activated simultaneously, (3) the difficulty of amending, generalizing, or creating scripts based on experience, and (4) general representation problems such as modeling physical objects, participant's point of view, and causal relationships. The compositional semantics approach is quite general, and has all the power of formal logic, but lacks the focusing mechanism provided by the frame structure. One way of obtaining this focusing mechanism is to couple the compositional semantics to an automatic planner, as is done in the KAMP system described in Box 6-5.

Pragmatics. To use language with the competence of a native speaker requires more than the description of syntactic, BOX 6-4 *(continued)*

user. LUNAR had an extensive grammar that covered a subset of English, and could handle some pronouns and definite determiners. Thus, it could establish a limited dialogue capability.

LIFER

The LIFER system [Hendrix 78] is a system for creating English language interfaces to other computer software, such as database management systems. The goal was to provide a systems designer who is not a linguist with the ability to tailor a natural language "front end" to an application. LIFER allows the systems designer to specify the nature of the processing to be carried out on the natural language inputs by writing pattern and response expressions. These can be thought of as more complex than, but similar to, the ELIZA patterns described previously. One of the useful features of LIFER is its ability to handle ellipsis. Thus in the series of questions "How old is John?" "How tall?" and "Mary?" the last two questions would be interpreted as "How tall is John?" and "How tall is Mary?"

When a given pattern is recognized by the parser, the associated expression is evaluated to produce the desired response. For example, a specification "HOW <ATTRIBUTE> IS <PERSON>" indicates that when an input sentence such as, "How old is John?" is entered, the system should identify "old" with <ATTRIBUTE> and "John" with <PERSON>. These "interpreted" words are then used in appropriate interactions with the application software. For example, a user can specify that the word "sum" as in "What is the sum of 3 and 4?" be used to call a summation function that uses 3 and 4 as arguments.

semantic, and discourse rules: human language behavior is part of a coherent plan of action directed toward satisfying a speaker's goals. Thus, pragmatics requires the use of reasoning and planning techniques, since the speaker must develop a plan of how to convert intent into a string of words, and conversely, the receiver must reason from the message to determine what that intent is.

The importance of considering the context of an utterance in deriving meaning is discussed by Searle in his classic book on speech acts [Searle 69]:

> The unit of linguistic communication is not, as has been generally been supposed, the symbol, word or sentence, . . . but rather the production or issuance of the symbol or word or sentence in the performance of the speech act. . . . More precisely, the production or issuance of a sentence token under certain conditions is a speech act, and speech acts . . . are the basic or minimal units of linguistic communication. A theory of language is part of a theory of action. . . .

Some of the problems that must be considered in pragmatics are how to deal with multiple sentences and extended discourse, and how to resolve references because such discourse analysis requires a model of what the participants know, believe, desire, and intend. (The referring problem was described in the "toy in the driveway" earlier in this chapter.)

Pragmatics in the CD Approach. When a script cannot make sense of a new input, possibly due to the fact that some additional pragmatic knowledge is required, the CD approach turns to *plans*, a set of actions and subgoals for attaining a

BOX 6-5 KAMP, A Program that Plans Utterances

The knowledge and modalities planner (KAMP) [Appelt 85] uses formal logic to plan utterances. Rather than go through the rather technical formalism, we will indicate the planning and reasoning used by describing a typical utterance problem solved by KAMP. The example shows the complex intertwining of reasoning and discourse operations. Consider a computer program, Rob, capable of performing speech acts but no other actions. A person, John, is to carry out the physical actions of repairing an air compressor. Rob is the expert and knows how to assemble the compressor, what tools are needed, and where the tools are located. Suppose that Rob wants the pump to be removed from its support. Rob reasons as follows:

1. For John to remove the pump he must unfasten the bolts attaching the pump to the platform.
2. To accomplish this, John must know what the right tool is, must have this tool, and must be next to the pump.
3. Rob assumes that John knows that the pump is attached to the platform so that it will not be necessary to tell him from what the pump must be unbolted. Thus, Rob starts to form the utterance, "Remove the pump," without including "from the platform."
4. A *critic* routine within the Rob program now indicates that it cannot be assumed that John knows the tool needed to carry out the removal. In addition, for John to have the wrench requires that he know where it is, and must go there and get it.
5. According to Rob's model of what John knows, John does not have this knowledge, so Rob must inform him of the need for the wrench and its location.
6. Rob now forms the complete utterance, "Remove the pump with the wrench in the toolbox."

It is difficult to imagine how people carry out these reasoning determinations so effortlessly and so rapidly!

---

goal. A knowledge of planning helps the program comprehend the motives of the actors. Once the plans and goals of a character in the text have been figured out, then guesses can be made concerning the intentions of an action in the unfolding story. From the CD point of view, to understand a narrative is to keep track of the goals of each of the characters in the text, and to interpret actions as a means of achieving these goals.

For example, consider the story, "John needed money for a down-payment on a house. He called his sister." To make sense of the story, we cannot expect to find a *paying for a house* script, and even if one did exist, it is not clear that the *sister* relation would be included. Thus, the system must have knowledge about the goal of raising money and plans for how this can be achieved. One plan might be *contact friends or relatives*. Since the sister is a relative, the connection between the original two sentences can now be made.

In this approach, an extensive set of plans must be stored in the database, and some technique for locating relevant plans must be provided. Notice that, as an attempt is made to deal with "deeper" meaning, a more and more sophisticated model of human actions is required.

Pragmatics in the Compositional Semantics Approach. In a formal approach to the planning and "decoding" of an utterance, logic is used to model the linguistic components of the discourse. An example of this approach is the knowledge and modalities planner (KAMP) [Appelt 85], which takes a set of axioms about the state of the world, the preconditions and effects of actions, the beliefs of different agents, and a description of a given agent's high-level goal, and produces an utterance plan that takes into account the abilities and beliefs of the other agents. Pieces of the utterance are constructed to supply information that the planner thinks the intended listener requires to understand the message and to carry out its intent. The linguistic actions are refined until an English sentence is completely specified. Box 6-5 indicates how the planning operation causes a sentence to be constructed based on the speaker's knowledge of the "local world" and a model of the recipient's knowledge and beliefs.

DISCUSSION

Language provides both a basis for social cooperation and a tool for thought. While many animal species can communicate, and the higher primates even seem capable of elementary forms of symbolic encoding of information, the full power of language use appears to be a distinguishing characteristic of the human species—perhaps its only distinguishing characteristic. The essential element of linguistic competence is a (shared) representation that is general enough to allow almost any situation of relevance (to the intercommunicating group) to be easily expressed, and is extendible, to allow one to deal with new concepts and situations.

While we easily recognize the importance of language for communication, how vital a role does it really play in our thinking and reasoning processes? If symbolic language was our only internal knowledge representation, then we would have to agree with the Whorfian hypothesis (see Chapter 3, "The Representation of Knowledge"). However, there is strong evidence to suggest that we have access to additional internal representations (e.g., iconic representations, see Chapters 8 and 9), and thus the role and importance of language in our thinking remains an open issue.

Where do we stand in terms of developing a machine that can use natural language at a human level of performance? In a very shallow sense, we have already reached this goal as noted in the case of the ELIZA program. In the deepest sense, it has been argued that we can never reach this goal since machines can never fully share human experience, and thus their conversation will always be distinguishable from that of a member of our culture. From that point of view the computer will always be an "alien intelligence," i.e., possibly intelligent, but lacking the "first-hand" experience with our culture to deal with linguistic situations like a native. But given these qualifications, we might still ask how far we have come in allowing a machine to carry on an intelligent conversation (i.e., the essence of the Turing test).

We saw that when we attempt to

devise programs capable of what appears to be advanced language understanding, we are faced with the Chinese room problem of determining what understanding really means—this issue is obviously very far from resolution. It further appears that the issue of language ability cannot be separated from that of intelligence. To have a machine participate in sophisticated discourse, we encounter the same problems faced in other AI domains: attaining human-level performance in reasoning, planning, and problem solving. The mechanical aspects of language production and understanding are at one end of a continuous scale; creativity and intelligence progressively impinge on linguistic ability as we move along this scale.

Despite advances in the field that have led to useful applications, particularly those involving human interaction in natural language with a retrieval database, there is still a long way to go. Important first steps have been taken for dealing with utterance involving multiple agents, but the knowledge and beliefs, and the plans and goals of all participants must be known for the analysis to proceed. Finally, the subtleties of understanding a joke, composing a sophisticated poem or story, or paraphrasing a complex body of text still elude us.

Appendixes

6-1

Representing Parsing Algorithms

This appendix briefly describes two approaches for describing and constructing parsing algorithms, the augmented transition network (ATN) and the chart grammar representations.

The Augmented Transition Network

The ATN network representation is a variant of the state transition diagram discussed in Chapter 3. It provides a convenient notation for specifying the operation of a given parser. As shown in Fig. 6-7, the ATN is a network of nodes and arcs, with symbols attached to the arcs that indicate what constituents must be recognized to traverse the arc. The network in the figure uses the constituents AUX (auxiliary verb), NP (noun phrase), and V(verb). Simple constituents such as noun and verb are identified by looking up the words in a dictionary, but auxiliary ATNs are used to recognize more complex sentence constituents such as NP and VP. Symbols on arcs show what constituent must be recognized to traverse the arc. Numbers indicate tests that must be satisfied to traverse the arc and/or the action to be performed. "Agreement" is with respect to the previous arc. Registers are provided to store

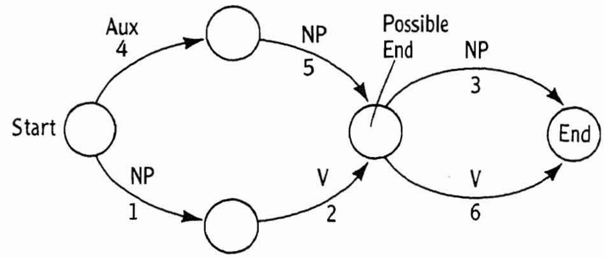


FIGURE 6-7 Example of an Augmented Transition Network.

Notes. Tests required to traverse arc and additional actions to be taken:

1 = Test: Constituent must be NP.
Action: Label constituent as "subject;" Label sentence "declarative."

2 = Test: Constituent must be V that agrees with subject.
Action: —

3 = Test: Previous constituent must be transitive verb; present constituent must be NP.
Action: Label present constituent "object."

4 = Test: Constituent must be AUX.
Action: Label constituent as V; label sentence as "question."

5 = Test: Constituent must be NP; constituent must agree with previous constituent V.
Action: Label constituent as "subject."

6 = Test: Constituent must be V; constituent must agree with previous constituent (subject).
Action: —

intermediate results. To parse a sentence we begin at the start node on the left and move through the network until we come to an end node. (There are several end nodes since a sentence could consist of NP + V, as in "The boy walked," or could have an additional NP as in, "The boy walked to school.")

The ATN of Fig. 6-7 can parse sentences such as "The boy walked home. Has the boy walked? The boy walked." For example, to parse "The boy walked home," we begin by determining from another ATN that "the boy" is a noun phrase (NP). Note 1 on that arc indicates that "the boy" should be labeled as "subject" and the sentence as "declarative." A dictionary then indicates that the next word "walked" is a verb, and note 2 on that arc indicates that if it agrees with the subject "the boy" then the arc can be traversed. Finally, an ATN will find that "home" is a noun phrase (NP), and note 3 indicates that if the verb "walked" is transitive, then "home" can be labeled as "object" and the arc can be traversed, completing the parsing.

The ATN formalism can concisely express a complex parsing procedure in an elegant form suitable for computer implementation. The disadvantage of the ATN is that it is difficult to modify large networks without causing unforeseen side effects, and the ATN formalism cannot conveniently describe efficient ways of searching for the required syntactic components.

Chart Parsing

The chart parser representation [Earley 70, Kay 73] is able to describe efficient ways of searching for relevant syntactic components. The chart uses edges in a graph to represent terminal symbols (words) and nonterminal symbols (such as NP). A sentence is parsed by constructing edges that span increasingly large sections of the original graph of terminal symbols (words). The computation is organized so that when a successful grouping of constituents is found (such as an article and a noun forming a noun phrase), these are retained for possible use when backtracking is required, while unsuccessful groupings are discarded. Parsing the sentence "The

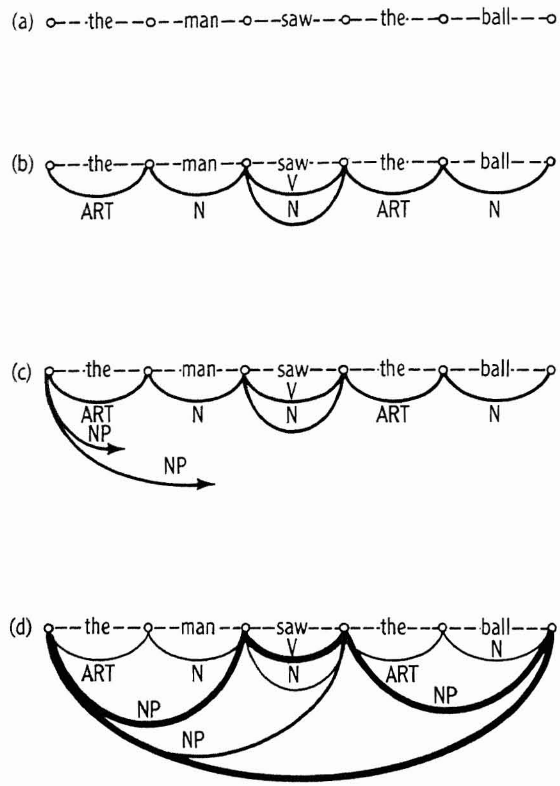


FIGURE 6-8 Example of Chart Parsing.

(a) Representing sentence using words as edges in a graph. (b) Assigning word classes. (c) Forming edges that span more than a single node. (d) Completed parse.

man saw the ball" would be represented as the following sequence of operations. We begin with the words of the sentence as edges in a graph (see Fig. 6-8[a]). The parsing of the sentence will consist of spanning increasingly large portions of the graph.

The first step is to assign the syntactic class of the word. If a word has multiple syntactic classes, multiple edges are used, as shown for the word "saw"(Fig. 6-8[b]). The system now tries to determine edges that subsume more than one edge, e.g., to form NP to span ART and NOUN. While in the process of trying to complete this spanning, an edge is either "active" or "complete." The chart parser uses rules concerning active and complete edges to control the spanning operation. In Fig. 6-8(c), two active NP's are in progress, one that will be complete after "the man," and the other after "the man saw," which, unless the lexicon indicated otherwise, could be considered by the system as a type of saw, e.g., the same noun-noun form as a "wood saw," or a "metal saw."

Eventually, a set of edges will satisfy the requirements for the sentence, as shown in Fig. 6-8(d).