Readings from

Scientific American

Human Communication

Language and Its Psychobiological Bases

With Introductions by
William S-Y. Wang
University of California, Berkeley

W. H. Freeman and Company
San Francisco
The Chinese Language

by William S-Y. Wang
February 1973

This melodious tongue is spoken by more people than any other. Although the Chinese system of writing is complex, the basic structure of the language is remarkably simple.

To people who are familiar only with the common European languages the Chinese language is strikingly different. Yet today Chinese is spoken by more people than any other language, and Chinese literature is the world’s oldest, spanning a period of 35 centuries. When we examine the structure of the Chinese language, we find that it is not conspicuously complex; indeed, in many ways it is simpler than the Western languages. But since Chinese does differ from the European languages in fundamental respects, some knowledge of its structure and historical development is indispensable to a general understanding of the nature of human language.

To the Western eye the writing system of the Chinese is altogether novel: instead of neat rows of simple alphabetic letters there are thousands of unique characters, many of which seem incredibly intricate. To the ear the language sounds rather melodious, perhaps a little like singing. When one peels below the surface, there are more surprises. The language has virtually no conjugation for its verbs and no declension for its nouns. The inevitable paradigms that Western schoolchildren have come to dread in their grammar books are totally absent.

The Antiquity of Chinese

Chinese is often termed a very old language. In a sense such a statement is misleading. All human languages go back to the dim uncertainty of prehistory, and at present we have no way of knowing whether or not they can all be traced back to the same root. Four thousand years ago the ancestors of the Chinese peoples spoke an early form of the Chinese language in much the same way that the ancestors of the English-speaking peoples were using an early form of the English language. Since almost nothing is known about the emergence of language in the human species, we are not in a position to say which of the world’s languages evolved earlier and which later. It is rather that in the course of history some languages have been renamed more often than others (as a result of events such as migration or conquest) and the newness of the names gives the illusion that the thing being named is new.

There is one sense, however, in which Chinese is a very old language. Sumerian is the only language we know of that has extant written materials that antedate Chinese ones. Sumerian cuneiform writing dates back some 5,000 years; the earliest Chinese writing in existence today dates back 3,500 years. But Sumerian and its derivative orthographies died out long before the beginning of the Christian Era. Chinese orthography has continued to this day, although there have been major stylistic changes.

The earliest Chinese writings are inscriptions on bone and tortoise shell. Most of the inscriptions are oracular, dealing with political or religious events or with the weather or warfare. Discovered toward the end of the 19th century in Chinese drugstores, where they were being sold as “dragon bones” for their medicinal value, the story of these inscriptions is a colorful chapter in the history of Chinese archaeology and philology. More than 100,000 inscribed pieces have now been found. Even though the total number of written characters on the pieces is more than a million, the number of different characters is small. The texts of the oracular inscriptions dealt with a very limited range of topics, and the same characters are repeated over and over again. Of the 2,000 to 3,000 characters on the shells and bones, about half can be read today.

Through the centuries Chinese characters have been preserved in many different mediums: metal vessels, stone drums, jade jewelry, coins, metal mirrors, bricks and tiles. The central line of development, however, has been the use of the brush on silk, bamboo, wood and ultimately on paper. A brush can produce variations in thickness whereas a stylus cannot. Such variations give the
MAP OF CHINA shows the distribution of the major dialects of the Chinese language. More than two-thirds of the Chinese population speak one of the Mandarin dialects, of which the speech of Peking is the standard. It is the dialects along the southern coast, however, that have been carried to many parts of the world by Chinese emigrants. There are also several non-Chinese linguistic stocks within China. The regions to the north and west are dominated by non-Chinese languages such as Mongolian and Tibetan.
writer a much greater artistic freedom in rendering his characters.

Some of the earliest written Chinese characters were pictographic. The character for "rain" was several columns of broken lines, and the one for "horse" looked like a horse, complete with mane and four legs [see illustration on page 67]. Pictographs, however, are only a minority in the total vocabulary of Chinese. Most of the words in the language cannot be suggested by a simple picture.

Calligraphy, the elegant rendering of characters, is a highly cultivated art form that has long been prized in Chinese culture, much as painting is valued in the Western world. For the Chinese there is a close relation between painting and calligraphy. Typically a silk scroll is covered with a picture and a few lines of characters, the two carefully balanced against each other. Because of their artistic values and their long history, Chinese characters have a much greater range of variability in their size and shape than the characters of any other writing system.

The Writing System

The Chinese writing system underwent major changes in 1956, when the government of the People's Republic of China decided to simplify the characters and also to adopt a system of spelling Chinese words in Latin letters. Both measures are intended to make the reading and writing of Chinese easier to learn, a crucial step in promoting linguistic unity and raising the standard of literacy in China.

In order to understand the nature of Chinese characters and their simplification, we must first examine their internal structure. Each character is made up of two types of smaller unit called the stroke and the radical. Roughly speaking, a stroke is a line, either straight or curved, that is completed every time the pen leaves the paper. For example, the character for "sun," which is pronounced ri, looks like:

It is built up of four strokes [see top illustration on page 57]. Both the order and the geometric position of the strokes are important. There are approximately 20 distinct strokes in the language, so that strokes are the closest counterparts to the 26 letters of the Latin alphabet. There is no counterpart of the radicals in the orthography of other languages. The traditional set consists of 214 radicals, and these radicals are found in almost all
There are a large number of characters in Chinese that are constructed on the phonetic-signific plan. Thus underlying many Chinese characters there is a phonetic principle. The average Chinese can often pronounce correctly a character he has never seen before simply by making a shrewd guess at its phonetic. For example, examine the following:

**銅**

The signific portion means "gold." The phonetic portion is pronounced "long" and is written:

```
龙
```

It gives its pronunciation to the character. On the other hand, there are characters that are not pronounced like their phonetic, often for reasons of historical sound change. The phonetic in the character "star" is:

```
生
```

The phonetic is pronounced "sheng," but the pronunciation for "star" is "xing.

As another example of a phonetic we can take the character for "horse." It is pronounced "ma" and is written:

```
马
```

When the phonetic for "horse" is combined with the signific for "woman," we have "ma," which means "mother":

```
妈
```

When the "horse" phonetic is combined with the signific for "jade," we have "ma," which means "agate":

```
玛
```

Combined with the signific for "insect," the meaning becomes "ant," and again it is pronounced "ma":

```
蚂
```

When there are two "mouth" significs hovering over the "horse" phonetic, the meaning becomes "to scold," and it is pronounced "ma":

```
骂
```

There would be no problem in agreeing that the character should be the name of a metal or metallic compound and should be pronounced like "long," even though such a character does not exist in the language.

Another fictitious character was sent to me recently by a friend as a riddle. The character has three components:

```
女上下
```

The first component means "woman." The second and third components also are independent characters; they are pronounced "shang" and "xia" and respectively mean "up" and "down.

For the character as a whole we settled on the meaning "elevator girl." However, there is no intuitive way of pronouncing the character, since the last two components do not constitute a phonetic. The solution of pronouncing it with two syllables, shang-xia, breaks a general rule of Chinese orthography, namely that for one character there is one syllable. Although reformers of the Chinese language have occasionally proposed polysyllabic characters, no such reform has ever been successful.

Chinese dictionaries and rhyme books may list tens of thousands of characters, but a knowledge of 4,000 to 7,000 characters is sufficient for, say, reading a newspaper. The form of the characters has been built up unsystematically through the centuries, and some are very intricate, requiring 30 or more strokes to write. The Chinese government's plan of simplification has gone a long way toward standardizing the form of the characters and reducing the average number of strokes per character. The net gain has been dramatic. A text written in simplified characters can contain fewer than half as many strokes as the same text written before 1956. This simplification makes the task of learning the written language considerably easier. An average of five or six strokes per character is not significantly different from the average of five or six letters per English word.

Since it is not the primary purpose of
the characters to represent sounds, the Chinese written language has been largely independent of the evolutionary changes that have taken place in the spoken language. This independence has made it possible for the written language to provide a literary continuity across thousands of years and to serve as a cohesive force binding the diverse cultures of China together.

The Evolution of the Language

The evolution of spoken Chinese, like the evolution of all other living languages, has been constant. Therefore many of the beautiful poems of the Tang dynasty of the seventh to 10th centuries no longer rhyme. If Confucius, who lived in the fifth century B.C., were to give a lecture anywhere in China today, he would not be understood. Within the large area of China dialects have evolved so far apart in their sounds that a man from Peking cannot be sure of being able to order a dinner in a Cantonese restaurant. Compared with the change in sounds, the written characters have changed little. Most of the characters Confucius used are still in books today, and many of these characters have their original meanings. The writing of Confucius is more intelligible to a modern Chinese than, say, a page of Beowulf is to an American. By the same token, although the Tang poems no longer rhyme, they are still enjoyed throughout China because their visual message remains the same. When the poems are read aloud by people in Peking, Shanghai or Canton, the poems sound altogether different because of the various dialects. It is rather like hearing “6 + 7 = 13” being read aloud in English, German and Norwegian. Even in Japan a Chinese with no knowledge of Japanese can manage to communicate reasonably well by writing. Chinese characters were also a significant medium of communication in Korea and in Vietnam. The independence of the characters from the spoken language has enabled them to serve as a core of culture in much of East Asia for many centuries.

A written Chinese character has a more direct connection with its meaning than a written word in English does. The sequence of letters spelling “horse” has meaning only through the mediation of the sounds they represent. The shape of the letters has no relation to the concept “horse.” Little would be changed if English-speaking peoples were to take up the Cyrillic alphabet and the sounds for “horse” were represented xopc. To a Chi-

<table>
<thead>
<tr>
<th>OLD</th>
<th>SIMPLIFIED</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUN (ㄖ)</td>
<td>日</td>
</tr>
<tr>
<td>STAR (xing)</td>
<td>星</td>
</tr>
<tr>
<td>MORNING SUN (long)</td>
<td>早</td>
</tr>
<tr>
<td>HORSE (馬)</td>
<td>马</td>
</tr>
<tr>
<td>MOTHER (妈)</td>
<td>妈</td>
</tr>
<tr>
<td>AGATE (玛)</td>
<td>玛</td>
</tr>
<tr>
<td>ANT (蚂蚁)</td>
<td>蚂</td>
</tr>
<tr>
<td>TO SCOLD (骂)</td>
<td>骂</td>
</tr>
</tbody>
</table>

Simplified characters were introduced in 1956 by the government of the People's Republic of China. Simple characters, such as the first two, were not affected. The remaining characters, somewhat more complex, were each reduced by six strokes.
Chinese the character for "horse" means horse with no mediation through the sound mā. The image is so vivid that one can almost sense an abstract figure galloping across the page:

The other major linguistic decision made by the Chinese government was to adopt a spelling system based on the Latin alphabet. This system is called Pinyin, which literally means "spell sound." All the Chinese words spelled out in Latin letters in this article are written in Pinyin. The government has been careful to point out that Pinyin is not intended to replace the characters but rather to serve as an aid in learning pronunciation. To discontinue the use of the Chinese characters would deprive coming generations of Chinese of a rich and meaningful cultural heritage.

To discuss the Pinyin system of spelling we need to examine the sound system of what is called standard Chinese, which is based on the Peking dialect and is now being taught in all parts of China. It is a straightforward system, simpler in many ways than the easy syllables of Spanish. There is really only one aspect unfamiliar to those who speak European languages: the tones. The musical quality of the spoken Chinese language is due to the fact that almost every syllable must carry one of four basic tones. These tones are indicated by diacritical marks over the vowels. The phenomenon of tones seems to be confined to Chinese and to some of the languages of Southeast Asia that have been heavily influenced by Chinese.

One of the earliest references in the literature to tones dates back to the sixth century. When the emperor of Li-äng asked one of his scholars, Zhōu Shē, what was meant by the four tones, Zhōu responded with an elegant illustration:

It means "The son of heaven is divine and wise." (Chinese emperors have traditionally been regarded as sons of heaven and divine, whether or not they were wise.) The beauty of Zhōu's response lies in the fact that the first word of his phrase illustrates the first tone in the speech of that time, the second word the second tone, and so on.

Essentially every Chinese syllable has a characteristic pitch pattern (tone). Changing the tone alters the meaning just as much as changing a consonant or a vowel in English changes the meaning. In English we use a rising pitch pattern for "John?" and a falling pitch pattern for "John!" The different tones convey different attitudes, but the meaning of the word remains the same. In Chinese, however, mā with a rising pitch pattern means "hemp" and mā with a falling pitch pattern means "to scold." The two meanings are no more related to each other than they would be if we were to change the vowel to get mī, which means "honey," or if we were to change the consonant to get pō, which means "to fear."

Standard Chinese has a total of four tones: rising, falling, level and dipping (see illustration at left). In addition to the tone every syllable must also have a nucleus to carry the tone, usually a vowel. The tone and the nucleus are the two obligatory components of the Chinese syllable. There are also three optional components of a syllable: the initial component, which is usually a consonant; the medial component, which is a glide, and the ending, which may be either a glide or a consonant from a restricted class. There are eight possible forms a syllable can take.
There are three optional components in a syllable: the initial, which is usually a consonant; the medial, which is a glide, and the ending, which may be either a glide or a consonant. All together there are eight possible forms that a Chinese syllable can take.

One striking feature of Chinese words in comparison with most European words is the lack of clusters of consonants before and after the nuclear vowel. When Western words with consonant clusters are represented in Chinese, they are typically broken up so that each consonant has its own syllable. “Marx” is conventionally rendered: 马克思

It is pronounced mǎ-kè-sī. The first character is the one for “horse,” which also happens to be a prevalent Chinese surname.

Although the Pinyin system and standard Chinese are taught in all parts of China, the languages of ethnic minorities are given full consideration. According to The Nationalities in China, a book published in Peking in 1961, there are about 30 million people belonging to minority groups. The minorities occupy about a third of the land of China, mostly in the west and northwest. Some of the groups are large: the Zhuangs are about 100,000 miles or so, the speech patterns of these villages are definitely of the rue dialect group, but they differ markedly from one another. We can be sure that significant differences have arisen in the past 100 years between the speech of the American Chinese and the speech of the source villages around Canton.

Word Formation

Every language has a stock of several thousand morphemes: the bearers of the basic semantic and grammatical content. An expression such as “can openers” comprises four morphemes: “can,” “open,” “-er” and “-s.” Some morphemes in English have more than one syllable (as in “open” and “adamant”), whereas others, such as the “-s” in “cats” and the “-t” in “slept,” are single consonants. In Chinese most morphemes are exactly one syllable long. The usual division of morphemes into three major categories—noun, verb and adjective—applies to Chinese quite well. Thus in Chinese shū ("book") is a noun, māi ("buy") is a verb and guī ("expensive") is an adjective. These morphemes are known as contentives, in that they carry independent semantic meanings.

There is also a class of morphemes called functives. They are usually attached to contentives to modify their meaning in systematic ways and to show the relations the contentives have to one another. For example, functives attached to the contentive “prove” change its meaning: “proves,” “proved,” “proving,” “disprove,” “unproved” and “proof.” Although conjunctional and declensional paradigms are important in English, they are negligible in Chinese. There is, however, a sizable amount of derivational morphology in Chinese, where nouns are derived from verbs by changing the tone. Thus shū is a verb meaning “to count,” and shā is a noun meaning “number.” The verb liân means “to connect,” and the noun liàn means “a chain.” Mō is “to grind,” and mō is “a grindstone.” To derive a noun in these cases one simply changes the syllable to a falling tone. A similar example from English is deriving a noun from a verb by devoicing the final consonant: “prove”—“proof,” “bathe”—“bath,” “house”—“houe.”

The device of derivation by tone change is no longer productive in standard Chinese, but it is suspected that tonal derivation was an important process in the earlier stages of development.
The computer translation of a Chinese sentence from a scientific text produces a reasonably accurate and understandable result. The string of Chinese characters is entered into the computer using a numeric code for each character. The position of each character in the sentence is also entered. The computer searches its memory for the meaning of each character and then performs syntactic analysis of the sentence. In converting the sentence into English, the computer makes permutations of the word order. The sentence means: "It may be obtained after bombarding the element with high-energy particles." Research into computer analysis of Chinese is being conducted by the author and his colleagues at the phonology laboratory of the University of California at Berkeley.
of the Chinese language. In the Cantonese dialect, however, the process is still very productive. The Cantonese use tone change for forming diminutives: "candy" from "sugar," "daughter" from "female" and so on.

A common derivational device in Chinese is reduplication. Applied to nouns, it carries the meaning "every." Hence although tén means "person," rénren means "every person," tán means "day" and tiándìn means "every day." Applied to verbs, it adds a transitory meaning to the action: kàn means "to look," whereas kànkan means "to take a look," zǒu means "to walk" and zǒusou means "to take a walk." Adverbs can be derived from adjectives by reduplication and the addition of a de suffix. Thus kuài is "quick" and kuāikuāide is "quickly," lān is "lazy" and lǎnlándé is "lazily."

The manner in which two-syllable adjectives reduplicate is different. Whereas a two-syllable verb, tāolùn ("to discuss"), reduplicates as tāolùntāolùn ("to discuss a little"), an adjective, say gāoxìng ("happy"), becomes gāogāoxìngxìngxìng ("happily"). A verb reduplicates by the entire word, but the adjective reduplicates in terms of its constituent syllables.

Another device for word formation in Chinese is the conjoining of antonyms. "Buy" and "sell" combine to form "business," mǎimá, "Long" and "short" combine to form "length," chángduǎn. The derived meaning is not always straightforward. For example, fán means "turned over" and zhèng means "right side up." Put together, fánzhèng means "in any case."

Classifiers are a linguistic feature peculiar to Chinese and its neighboring languages. Articles, numerals and other such modifiers cannot directly precede their associated noun; there has to be an intervening classifier, which usually has negligible semantic content. In Chinese one cannot say sān shū ("three books") or nèi mào ("that cat"). One has to say sān běn shū ("three piece book") or nèi zhì mào ("that piece cat"). The terms běn and zhì here are translated as "piece" for the lack of a better counterpart in English. Such classifiers are absolutely necessary in Chinese expressions. This feature has been carried over into many pidgin and Creole languages based on Chinese, in which "three bananas" or "this man" are rendered as "three piece banana" or "this fellow man."

**Sentence Formation**

The basic sentence in Chinese has the order subject-verb-object, as in English. Thus the sentence wōmén chī fēi is word for word "We eat chicken." There is a tendency in Chinese to delete either the subject or the object. Hence wōmén chī ("We eat") or fēi fēi ("Eat chicken") are both common sentences. To focus attention on the object, the Chinese speaker will move it to the beginning of the sentence. Fēi wōmén chī, for instance, means "We eat chicken," but it is normally used to contrast with something else we do not eat. Suppose subject deletion and moving the object to the initial position both operated on the sentence wōmén chī fēi. First we front the object and get fēi wōmén chī, and then we delete the subject and get fēi fēi. The resulting sentence would mean something like "Chicken A eats," and the identity of A normally would be clear from the context. But fēi fēi of course also means something like "The chicken eats," where fēi is the subject of the verb. In other words, fēi fēi is an ambiguous sentence, its ambiguity arising from object fronting and subject deletion.

There is no evidence that Chinese allows either more ambiguity than English or less. In an English sentence such as "It is too hot to eat," the "it" can refer to the weather, to the food or to the animal that is doing the eating. Moreover, "hot" could mean "high in heat content" (which in Chinese is tāng) or it could mean "spicy" (which in Chinese is là).

The Chinese language as an object of study goes back as far as the beginning
of the Christian Era. At that time the Chinese had already produced sophisticated works in dialectology and in semantic classification. The foundations for the historical study of Chinese sounds were laid during the 17th and 18th centuries by the great scholars of the Qing dynasty, and it is on their shoulders that modern Chinese linguists stand.

The key to how a language sounded centuries ago lies in how it sounds today. The basic method is to compare the pronunciation of morphemes in contemporary dialects and to infer what their ancestral pronunciations might have been. The inference is not just a guess but is made on the basis of documentary evidence and knowledge of the general linguistic principles underlying sound change.

Given the non-phonetic nature of the Chinese writing system, it may seem an impossible task to reconstruct how the language was spoken many centuries ago. One might think that the sounds of alphabetic languages, such as Old Church Slavic or Sanskrit, would be much easier to reconstruct. Actually it is not much easier, because there is no direct way to determine how a letter was pronounced. With alphabetic languages the phonetic values must also be arrived at by inference.

Chinese has the great advantage of an abundance of ancient writings that reach back continuously in time further than the literature of any other language in the world. The fact that the form of Chinese characters is often not much influenced by changes in pronunciation is quite a convenience in helping linguists to determine which morphemes are etymologically related.

The study of Chinese dialects has been hampered in the past by an overabundance of data, which tended to make research procedures cumbersome and time-consuming. The advent of large computers has facilitated the manipulation of the data. Chin-Chuan Cheng and I, with the help of several colleagues, have developed a dialect dictionary on computer, which we call doc. The program is in operation at the computer centers of the University of California at Berkeley and of the University of Illinois. It incorporates the pronunciation of more than 2,000 morphemes in each of 20 Chinese dialects.

Evidence is accumulating from our work with doc that changes in language proceed in ways that are essentially parallel to biological evolution, as Charles Darwin noted in *The Descent of Man*. In both cases the mechanism of change resides in variation. When two or more variants appear, the rival forms compete for survival. For instance, in American speech the vowel in "room" varies between that of "pool" and that of "put," and the s in "disobey" varies between an s and a z. The major selective force, which is constant across time as well as across languages, is the ease with which the forms can be pronounced and properly perceived. This selective force determines what forms of speech will survive.

With the aid of computer programs such as doc we have been able to run through large pools of data to locate ongoing sound changes of special interest. Some changes are just beginning, some are in midstream and some are ending. In the Chaozhou dialect, for example, there is a change from one tone to another that has so far affected about half of the vocabulary. By having access to a large number of well-defined types of sound change, we are now in a position to study much more effectively the selective force that determines the direction of language evolution.

Another aspect of the Chinese language that has been subjected to intensive computer study is its morphology and syntax. With the collaboration of Stephen W. Chan, Benjamin K. Tsou and others, we have developed a machine dictionary with more than 70,000 entries, together with the necessary programs for translation from Chinese into English. A good measure of how well we understand the structure of a language is how well we can break it apart in a way that is suitable for translation by machine. Although we are an indefinite distance away from being able to translate a Tang poem into English without losing its exquisite sensitivity, we can do a reasonably accurate job with scientific texts [see illustration on page 60].

Now that relations between the People's Republic of China and Western nations are becoming more normal, interest in the Chinese language is increasing at an accelerating tempo. Considering that the Chinese language has the largest number of speakers in the world and the greatest time depth in its literature, this interest is long overdue. With the increase in interest we may look forward to deeper probing into the history and the structure of the language, and into the influence the language has had on the cultural and intellectual development of the Chinese people. These studies will surely lead to a better general understanding of the nature of human language.