

1

Introduction

1.1 The varieties of Chinese

Few language names are as all-encompassing as that of Chinese. It is made to serve at once for the archaic inscriptions of the oracle bones, the literary language of the Zhou dynasty sages, the language of Tang and Song poetry and the early vernacular language of the classical novels, as well as the modern language in both its standard and dialectal forms. And this list is by no means exhaustive. This creates a certain awkwardness when one wants to speak about one of the varieties of Chinese; if ambiguity is to be avoided, it is necessary to employ a complex designation of some sort – Classical Chinese, Literary Chinese, Middle or Ancient Chinese, Early Vernacular Chinese or, in the case of a modern dialect, a geographical designation like Shànghǎi or Fúzhōu dialect. In this book we will examine Chinese in its many diverse aspects and in the process hope to clarify some of their differences and interrelations.

Why have so many disparate historical stages and geographical variants of a linguistic continuum like this been subsumed under a single name? After all, the modern Chinese dialects are really more like a family of languages, and the Chinese of the first millennium BC is at least as different from the modern standard language as Latin is from Italian or French. The explanation is to be found in the profound unity of Chinese culture that has been transmitted in an unbroken line beginning from the third millennium BC and continuing down to the present day. Even in periods of political disunity at various times in the past, the ideal of a single, culturally unified Chinese empire has never been forgotten. The Chinese language, especially in its written form, has always been one of the most powerful symbols of this cultural unity. The aptness of language as a symbol of cultural and even political unity was facilitated by the use of a script that for all practical purposes was independent of any particular phonetic manifestation of their language, allowing the Chinese to look upon the Chinese language as being more uniform and unchanging than it actually was. Such a view was no doubt also reinforced by the use of a literary language which changed but

little from century to century and from dynasty to dynasty. When one adds to these considerations the fact that the Chinese throughout most of their history have been conspicuously uninterested in their spoken language, precisely the area where linguistic variation would have been most evident, it is not hard to see why they considered so many linguistically disparate forms to be a single language. In our own nomenclature, we in the West have simply adopted this Chinese idea.

There can be little doubt concerning cultural continuity in Chinese history, but can a similar degree of cultural cohesion be observed over the vast territory of modern China? In a practical sense, the Chinese dialects differ from one another quite dramatically. A speaker of the Peking dialect can no more understand a person speaking Cantonese than an Englishman can understand an Austrian when each employs his native language. Nevertheless, the Cantonese speaker feels a much closer cultural affinity to the Peking native than an Englishman does to an Austrian. In part this can be explained historically; throughout most of Chinese history Canton and Peking have belonged to the same political entity, and have always used the same written language. England and Austria, on the other hand, have not been united politically since the days of the Roman empire. Moreover, after the Renaissance and Reformation, England and Austria lost many of the linguistic, cultural and religious ties which existed in Western Christendom in the Middle Ages; perhaps the most important factor in creating such different cultural entities in Europe was the abandonment of medieval Latin as a common literary language and the adoption of different local languages for political, religious and literary purposes. China did not experience a comparable break with the past until the present century; and when the Chinese began to abandon their old common literary language, they did not sacrifice the country's linguistic unity: they simply replaced the old literary language with a new unified standard based on the dialect of the capital.

There is another factor which may be viewed as a cause of China's continuing linguistic unity. In Europe, since at least the end of the first millennium AD, a large number of literary languages based on local vernaculars arose and in subsequent centuries many of these early written languages became the bases of several of the most important national languages of modern times. Nothing like this seems to have happened in China. The vernacular literary language which began to evolve in the Tang dynasty (AD 618–907) and gradually developed into the vehicle of a flourishing vernacular literature in the Song (AD 960–1279) and Yuan (AD 1271–1368) dynasties was based on a northern variety of spoken Chinese. As far as is known, no comparable written language based on a non-standard dialect (especially the very aberrant dialects of southeastern China) ever

developed. Hence, dialectal forms of Chinese generally lack a historical basis on which to build a literary language to compete with the national standard. The present-day standard written language, on the other hand, is a direct continuation and development of the literary vernacular of Song and Yuan times.

This is not to say that non-standard Chinese dialects have never been written; they have. But they have always been the vehicle for certain types of local literature which traditionally have not enjoyed high repute among the Chinese educated classes. There seems to be no case where a non-standard dialect has ever been employed in a written form as the language of administration or even of commerce. Even in periods when China has been disunited, there has never been an attempt to set up a regional literary language based on one of the local dialects. Beginning in the middle of the nineteenth century, Christian missionaries began to write certain local dialects, generally in romanized form, as part of their effort to evangelize the common people of China. A few of these written dialects, especially that of Amoy, enjoyed a certain success among converted Christians, but they have generally been viewed with hostility by Chinese authorities, and are little used nowadays.

Practically speaking, there have been only two forms of written Chinese throughout most of Chinese history: the classical literary language, based on the prose classics of the Late Zhou and Han periods, and the vernacular literary language which first arose during the Tang dynasty. In both cases, the prestige of the established norms was apparently sufficient to block the development of any dialectal or regional competitors.

The notion of a single Chinese language existing in a great number of forms, both written and spoken, has important implications for Chinese linguistics. For one thing, it is very difficult to draw sharp boundaries between the different varieties of the language: the colloquial is permeated with elements taken from the literary language, and many texts written in the literary language clearly betray the influence of the contemporary vernaculars. The current standard language is an amalgam of the popular Peking dialect, loans from the literary language, and even features of other present-day dialects. The Chinese are of course aware of the differences between the modern vernacular-based standard and the older literary language; but they probably do not think of them as two utterly different languages, as the typical foreign observer is prone to do. The poetry of the Tang dynasty can be read and enjoyed by most university graduates in China; the same can scarcely be said of *Beowulf* or even of Chaucer in the English-speaking world.

To study any one of the many different forms of Chinese, it is necessary to possess at least some familiarity with other important varieties of Chinese. In

surveying dialects, for example, it is essential to have a good knowledge of the national standard language and, in some cases, of the provincial or local standard as well; lacking such knowledge, the dialectologist will be hard-pressed to distinguish what is genuinely local in the dialect he is studying and what are importations from some type of standard language. The study of older texts is complicated by similar factors; the researcher must be able to distinguish traditional literary elements from vernacular intrusions. Neither of these tasks is entirely straightforward and easy to carry out. The linguist who deals with Chinese is nonetheless constantly faced with the problem of distinguishing different varieties and strains in the particular language he is dealing with.

The history of the Chinese language can be traced back to approximately the middle of the second millennium BC, to the period known historically as the Shang dynasty (sixteenth to eleventh century BC). The texts which survive from that time are divinatory texts inscribed on bone and shell; they tend to be short and formulaic, but there is little doubt that they are written in a language closely related to the far more abundant written records of the succeeding Zhou dynasty, (Zhōu 1956, 129). The Zhou dynasty (eleventh to third century BC) language is preserved in bronze inscriptions and in copious literary texts which have survived down to the present day; the language of these texts is unquestionably ancestral to all later stages of Chinese, including the modern dialects.

At this early period the Chinese language was spoken in a much smaller territory than it is today. Its center was the Yellow River Plain, but even during the Zhou dynasty it had doubtless already begun to spread to some peripheral areas. In subsequent centuries the Chinese (by which I mean Chinese-speaking peoples) were to spread outward, gradually overcoming and assimilating their ethnic neighbors until they would finally occupy the vast territory that they do at the present time.

The language of such well-known Zhou texts as the *Analects* of Confucius and *Mencius* was probably not very different from cultured contemporary speech. The strict distinction between the written language and the vernacular was something that was to develop in later times. The schism between the written literary medium and the spoken vernacular began to develop in the Han dynasty (206 BC–AD 220), but the differences between the two forms of Chinese may still not have been very great. Naturally the gap between what was written and what was spoken increased in direct proportion to the distance from the formative period of the written language: by the Qing dynasty (AD 1644–1911) the difference between the literary language and the popular vernaculars was very great indeed.

Since literary Chinese has been modeled on the language of the first millennium BC, it is very difficult to trace the development of vernacular Chinese

1.1 *The varieties of Chinese*

5

in subsequent centuries. Vernacular elements do appear in written sources, to be sure, but they are generally sporadic intrusions into otherwise literary texts. These vernacular elements are chiefly to be found in texts of popular origin such as Buddhist tales, anecdotal collections, and folk poetry. A full-blown vernacular literature began to appear in the Song dynasty and has developed in a direct line down to modern times: it is in this “old vernacular” that most of the great fictional prose of China is written.

Regional dialects have surely existed in China from the most ancient period, but the written record tells us very little about them. The *Fāngyán* of Yáng Xióng (53 BC–AD 18) is an early record of dialect words; unfortunately the Chinese script does not allow us to form any definite idea of how these words were pronounced. Aside from this work, however, there is very little information to be found concerning Chinese dialects right up to modern times. What little there is is sketchy and rarely interpretable in phonetic terms. In the late Ming dynasty (AD 1368–1644) a few dialect dictionaries appeared; but they were chiefly guides to the proper pronunciation of characters found in literary texts, and are rather poor sources for the study of dialectal vocabulary and of no help at all in the study of syntax. It was only with the arrival of Western missionaries and officials that more detailed descriptions of dialects began to appear. These works, valuable as they are, are marred by the fact that their authors were rarely trained linguists. The scientific study of Chinese dialects began with the work of Bernhard Karlgren and Y. R. Chao in the early part of the present century. Karlgren was interested in Chinese dialects only to the degree that they reflected the pronunciation of what he called Ancient Chinese; Chao, on the other hand, was interested in the dialects as subjects worthy of study in themselves. He personally conducted the first scientific survey of the Wú-speaking region, and planned surveys of the entire country. Unfortunately, because of the war with Japan in the 1930s, only a few of these surveys were ever actually completed. The work of Karlgren and Chao nonetheless provided a solid foundation for later work, much of which has been carried out by their students.

The lack of dialectal texts in earlier times indicates that, throughout most of her history, China has had a standard language; this is clearly the case for the written language, but there are numerous indications that it was also the case for the spoken language as well. It especially stands to reason that, during the great periods of imperial consolidation and bureaucratic centralization, there must have been a common spoken medium or *koine* corresponding to the common written language in which the business of government was carried on. The present standard language has its roots in just such a *koine*; it arose in the foreign dynasties of Liao (AD 916–1125), Jin (AD 1115–1234) and Yuan (AD

1271–1368), all of which established their capitals in the region of present-day Peking.

The present work will examine the Chinese language in its many historical and contemporary variants. It is my belief that none of these variants, be it ancient or modern, standard or regional, can be properly understood without reference to the Chinese language viewed as a whole. It is evident to most that the modern language along with all its dialects is fully intelligible only when examined in a historical perspective. I would also contend that earlier stages of Chinese as revealed in written records can often be better understood when examined in the light of modern spoken forms of the language.

1.2 China's linguistic neighbors

In the remainder of this chapter, Chinese will be examined in relation to the languages which are to be found on her periphery. In doing this I will limit myself to those language groups which are actually contiguous to Chinese at the present time; in this way the necessity of discussing the Indo-European languages of India and Iran is eliminated.

To the north of the Chinese-speaking area are found various languages belonging to the Altaic family. This family is composed of several subgroups: Turkic, Mongolian, Tungusic and, according to some, Japanese and Korean (Miller 1971). All these languages are non-tonal and polysyllabic, typologically very unlike Chinese. Grammatical relationships are mainly shown by a set of suffixes. The classic Altaic subgroups – Turkic, Mongolian and Tungusic – all display vowel harmony to varying degrees, and there appear to be vestiges of this in Japanese and Korean also. All these languages employ Subject–Object–Verb word order; indeed, the syntax of all the above-mentioned language groups is so similar that it is often possible to translate from one to another word for word with few, if any, changes of word order. Comparative work on Turkic, Mongolian and Tungusic is well advanced (Ramstedt 1952; 1957; Poppe 1960). A few European Turkologists remain sceptical about the Altaic theory, but among Mongolian and Tungusic specialists, the Altaic affinity is routinely accepted.

The highly complex and diverse Tibeto-Burman family of languages lies to the west and southwest of China. This family is comprised of scores if not hundreds of languages, many of them spoken by small tribal groups. Tibetan and Burmese, the two languages from which the group takes its name, have long written traditions, but the vast majority of these languages have either never been written or have been reduced to written form only in modern times. While the genetic affiliation of these languages to one another is unquestionable, comparative work is still in its infancy. There is no really adequate comparative study of the

group as a whole, but work on some of the subgroups has progressed to a high level of sophistication (Matisoff 1970; 1972).

Chinese is considered by most Tibeto-Burmanists and specialists in Chinese to be a rather distant cousin of this group. We will examine this problem in more detail below.

The region south of the Chinese-speaking area is home to several different ethnic groups possessing their own distinctive languages. The Miao-Yao peoples are the northernmost of these groups; at the present time they are found chiefly in western and southern Húnán, northern Guǎngxī, Guǎngdōng, and in a few areas in Yúnnán and Sīchuān. Isolated groups of Miaos can also be found in the extreme southwestern corner of Húběi, and on Hǎinán island. Both Miao and Yao settlements are also encountered in Vietnam, Laos and Thailand, but these are due to rather recent migrations.

While the relationship of the numerous Miao and Yao dialects to one another is not questioned, the wider genetic relationships of Miao-Yao are still very much a matter of controversy. Comparative work on the family by Chang (1947; 1953; 1966), Downer (1963; 1967), Haudricourt (1954c; 1966), and Purnell (1970) will no doubt aid in solving this problem.

The Tai languages taken in the narrow sense form a closely knit group of languages and dialects spoken in the Guǎngxī Zhuang Autonomous Region and Yúnnán, as well as in Vietnam, Laos, Thailand and Burma. This language family is probably the best studied in Southeast Asia from a comparative point of view. F. K. Li's authoritative *Handbook of comparative Tai* (1977) is a crystallization of a lifetime of work on this topic. There is little doubt that Tai is also related to the Kam-Sui and Bê languages, and (perhaps somewhat more distantly) to the poorly known Kelao, Lakkia and Li languages; the still more obscure Lati and Laqua languages of the China–Vietnam border region are also doubtless related in some way.

In 1942 Paul Benedict proposed that Tai and what he termed Kadai (a cover term for the more distant relatives of Tai: Li, Lati, Laqua, and Kelao, etc.) are related genetically to Austronesian, and that the hypothesis of a Sino-Tai relationship, widely held at that time, was invalid. Although the evidence which Benedict gives to support his theory is in some cases very suggestive, his theory has by no means gained universal nor even general acceptance, especially among specialists in Tai and Chinese linguistics. Now that comparative work on both Tai and Austronesian has progressed far beyond what it was at the time of Benedict's original suggestion, the question needs to be re-examined in a careful and systematic way.

Austroasiatic languages are spoken over a vast geographic range: the Munda

languages in northwestern India, Khasi in Assam, Palaung-Wa and Mon in Burma, the Mon-Khmer languages in Indo-China, Vietnamese and Muong in Vietnam. At the present time only three languages of this group are indigenous to China, Va, Blang and Benglong, all spoken along the Burmese border in Yúnnán province. Evidence that Austroasiatic languages were once spoken much more widely in China, and may in fact have influenced the development of Chinese in some ways, will be examined below.

Languages of the extremely widespread Austronesian family, which includes such languages as Indonesian, Samoan and Hawaiian, are spoken in the mountainous areas of Tái-wān by a number of different tribal groupings. Theories that Austronesian languages were once spoken on the mainland of China are sometimes encountered (Solheim 1964); but such theories are almost always based on archeological or anthropological arguments, and cannot be verified linguistically.

1.3 Typological characteristics of Chinese

Languages basically can be compared in two ways. Their lexicons and morphologies can be compared with a view to determining whether they have a common origin; to prove this sort of genetic relationship, it is essential to show a pattern of recurring regular phonological correspondences in the lexicon. It is also possible to compare languages in order to classify them into phonological, morphological and syntactic types. When this is done, the groupings arrived at frequently cross genetic lines. For example, Modern Chinese and Trique (a Mexican Indian language) are both typologically tonal languages, but it is highly unlikely that a genetic relationship between them could be demonstrated.

In this section some of the general typological features that have often been used to characterize Chinese and other languages in East Asia will be considered. In every case, Chinese shares the features in question with other neighboring languages; this no doubt is in part due to the mutual influence that these languages have exercised on one another historically, and does not necessarily point to any degree of genetic affiliation.

Many of the languages of East Asia, including Chinese, are monosyllabic. This is generally taken to mean that in these languages morphemes are by and large represented by single syllables. Actually there is probably no language in which all morphemes are monosyllables; but in the type of language that we are referring to here, the vast majority of morphemes do in fact consist of single syllables. This is the case with Chinese at all stages of its development. In Modern Chinese there are many polysyllabic *words*, but these almost always consist of strings of monosyllabic morphemes: a word like *diàn-huà* 'telephone' is made up of two morphemes meaning 'electric-speech' respectively: genuinely polysyllabic mor-

1.3 Typological characteristics

9

phemes like *zhīzhu* 'spider', *dāla* 'hang down', *gēda* 'lump' are decidedly in the minority.

Chinese, along with many other contiguous languages, is a tone language. In languages of this type, each syllable is characterized by a fixed pitch pattern; such pitch movement may be level (neither rising nor falling), or it may be a contour (rising, falling, or some combination of the two). Tones are phonemic, in that they may serve to differentiate meaning just like consonantal and vocalic segments. Standard Chinese has four tones: the first has a high-level pitch, the second is high rising, the third low falling–rising, and the fourth begins at high pitch and falls abruptly; examples of tones are given below:

Tone 1	Tone 2	Tone 3	Tone 4
bā 'eight'	bá 'pull out'	bǎ 'grasp'	bà 'dam'
xī 'tin'	xí 'mat'	xǐ 'wash'	xì 'opera'
liū 'slide'	liú 'flow'	liǔ 'willow'	liù 'six'

Chinese dialects and the other tonal languages of Asia vary greatly in the number of distinct tonal categories which they possess: complicated systems have seven, eight or (rarely) more tones; four or five tones would be considered a fairly simple tonal system. A very few Chinese dialects have only three tones.

Tonal languages are not only extremely widespread throughout Southeast Asia and China, they also show remarkable similarities to one another. Gordon Downer (1963) has shown that the tonal systems of Chinese, Miao-Yao and Tai have developed in almost identical fashion. To these groups we could also add Viet-Muong (Thompson 1976) and Lolo-Burmese (Matisoff 1972); in all these language families, an originally rather small number of tonal contrasts (usually four) has been increased under the influence of features found at the beginning of the syllable, chiefly voicing; occasionally other features of syllable onset, such as aspiration and glottalization, have also led to the formation of new tonal categories. In subsequent chapters we will examine how this has taken place in a number of Chinese dialects. Since at least the time of the *Qìyàn* dictionary (AD 601), the source on which the reconstruction of Middle Chinese is based, Chinese has lacked consonant clusters; but there is strong evidence that in Old Chinese (first millennium BC) a variety of consonant clusters could occur at the beginning of the syllable and perhaps, in a much more limited way, at the end of the syllable as well. The simplification and eventual loss of all consonantal clusters appears to be a tendency affecting most, if not all, East Asian languages. Mon-Khmer, Tai, Miao-Yao, and Lolo-Burmese all exhibit this tendency, although at least some of the languages of each of these groups still keep

a certain number of initial clusters. Chinese, taken as a whole, has progressed furthest in this regard. It may be that the age-old contacts between Chinese and Altaic languages spoken to the north of Chinese territory have influenced Chinese along these lines, since Altaic, as far as can be ascertained, has never tolerated any sort of initial consonant combination at the beginning of words (Poppe 1960).

In traditional typological schemes, Chinese has been considered the pre-eminent example of an isolating or analytic language. By this it is meant that in Chinese the word was by and large coterminous with the morpheme, and that grammatical relationships were shown either by word order or by the use of independent grammatical particles, rather than by affixes or by internal changes in the word itself. This is a reasonably accurate way of describing Chinese at all of its historical stages. Modern Chinese dialects have developed a very small number of quasi-suffixes which function as grammatical determinatives, but grammatical relationships are still mostly shown by word order and by particles. In contrast to the complex suffixing languages of the Altaic family, Chinese indeed appears very analytic, and in this it resembles the languages of Southeast Asia, especially those of the Tai, Miao-Yao, Lolo-Burmese, and Viet-Muong groups. Even the Mon-Khmer languages have only very poorly developed affix systems, and may be considered at least mildly analytic.

Another widespread syntactic feature shared by Chinese and the languages of Southeast Asia is the use of measures (or classifiers) with numerals and determinatives. In these languages a numeral or determinative cannot be used alone with a noun: 'three men' in Chinese is *sān-ge rén* where *sān* is the numeral, *ge* the measure, and *rén* the word for 'man'. The order of these elements varies considerably from language to language, but the principle is much the same. Measures have spread to Japanese and Korean under Chinese influence, but in both languages there are still cases in which measures are not obligatory. Even in Chinese it appears that the obligatory use of measures with numerals and determinatives is a relatively late development; the classical language permitted both numerals and determinatives directly before nouns without any intervening element. Unfortunately, we know little about the historical development of this syntactic feature in the other languages where it is found at present.

Word order has come to be considered an important index of typological classification. The relative positions of subject (S), verb (V) and object (O) are a primary feature of any language; in the East Asian area, it neatly divides all languages into two types – those in which the object precedes the verb, and those in which it follows. The Altaic languages and most of the Tibeto-Burman languages have SOV order; Chinese, Tai, Miao-Yao, Viet-Muong, and Mon-Khmer