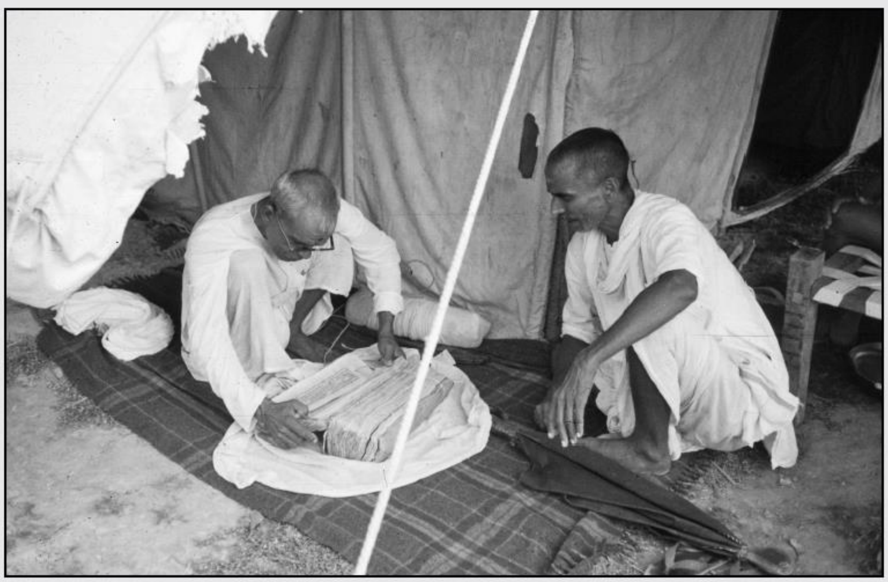
2

Tongues, Texts, and Scripts



If any evidence were needed of Asia's tremendous diversity, you need look no further than language. There are over 2,000 distinct languages spoken in East, South, and Southeast Asia. Language both creates communities through mutual intelligibility—a group in a room with a light on—and divides people, possibly even close neighbors, across language borders in the darkness of nonintelligibility. Language may, in fact, be the most severe of cultural boundaries, for when you cannot understand other people it is easy to suspect them of any number of other contrasts with yourself: *they* speak gibberish; *they* are barbarians (heathens, cannibals); *they* are uncivilized (uncultured, primitive, savage); *they* are irrational; *they* should learn from us. On the other hand, language creates communities of mutual comprehension and becomes the base for other markers and reinforcers of community. We who speak a common language also share a common culture, have a sense of common identity, surely must have a common history and common ancestors, and are a common "people," a nationality, an ethnicity. Many of these commonalities are actually fictions—but important fictions in the creation and maintenance of social unity.

For instance, in the present age when so many national borders enclose people with no common language, there may be a desperate search for one. In India, where there are over 150 distinct languages, the effort to create a common one has been intense and controversial. The dominant language, Hindi (which is the fourth largest linguistic community in the world after Mandarin, English, and Spanish), has been hotly resisted in South India where speakers of Dravidian languages resist imposition of a northern language. If there is a neutral common language, it is English, although the irony is that this is the language of the former colonial power. And, of course, a great number of Indians never learn this "common" language.

China, too, is a country of linguistic diversity, but it deals with the problem somewhat differently. It simply declares that all ethnic Han Chinese (92 percent of the population) speak "dialects" of Chinese, even though these "dialects" are mutually unintelligible, and linguists consider them separate languages. The "Chinese" language is actually a language family equivalent to the "Romance Family" of Italian, French, Spanish, and Portuguese. It helps, of course, that in written form, speakers of these separate languages of Chinese can actually read the same texts (see later in this chapter). It is as if all speakers of the various European languages had a single written language that everyone could understand.

Chapter opener photo: A Brahman consults the genealogist about a potential marriage for his daughter. Genealogies 24 generations long are maintained in traditional style texts in this part of India.

On the other hand, Lao and Thai are mutually understandable and are therefore technically dialects, but since they are the national languages of Laos and Thailand, they are considered by their own national speakers to be two separate languages. This is one of those cases about which it has been said that a language is a dialect with its own army and navy.

There are many reasons for the study of language. You might want to study a particular language so you can speak it: that is, so you can go into the lighted room and converse with other speakers of that language. That may seem like the main point of language study. But that will not be the point of this chapter, nor could it be. Why else, then? In this book we are taking a view of Asia that seeks interconnections among people, one that is not content with a "Holistic Asia" view, or with the present nation-state configuration, or with any simple view of the "great traditions" as bounded and eternal. Rather, we are attempting to order the diversity. The diversity of Asia's languages is not truly chaotic, though occasionally it might seem so. Although in one sense the languages appear as so many separate lighted rooms, in another sense we are coming to know a great deal about the relationships between these languages and about the larger house, or family, of which each one is a part. The study of language families often uncovers startling connections—you might be surprised who your "relatives" are—but beyond this, historical linguistics is a key tool of prehistorical research, along with archaeology, paleontology, and biogenetic research.

Voices from the Past

Making Family Connections: The Indo-Europeans

In 1784 a young Englishman named William Jones rented a thatched bungalow 60 miles upriver from Calcutta and began taking lessons in Sanskrit. The learned Brahmans at first refused to teach their sacred language to a foreigner, since not even their own women and lower castes were allowed to speak or hear it, but Jones managed to find a non-Brahman who would teach him. Only one other Englishman had learned the language before him. It proved to be an exceedingly difficult language even for a linguistic whiz kid like Jones. Before graduating from university he had already mastered French, Greek, Latin, Persian, and Arabic. Within a few months of beginning his lessons, he began to notice remarkable similarities between Sanskrit, which he was learning, and Greek and Latin, which he already knew. They even look suspiciously similar to English speakers. The Sanskrit word for mother was *matr*, for mouse was *mus*, for name was *nama*, for two was *dva*, for three was *tryas*, and so on. In grammar, too, the similarities were impossible to miss. Clearly he was on to something important. Two years later, he announced to the Asiatic Society:

> The Sanskrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either; yet bearing to both of them a stronger affin-

> ity, both in the roots of verbs and the forms of grammar, than can possibly have been produced by accident; so strong, indeed, that no philologer could examine them all without believing them to have sprung from some common source, which perhaps no longer exists. There is similar reason, though not quite so forcible, for supposing that both the Gothick [i.e., Germanic] and Celtick, though blended with a different idiom, had the same origin with Sanskrit; and the old Persian might be added to the same family. (Franklin 1995:361)

Prior to this time, European scholars who noticed similarities between, say, English and German or English and French, assumed it was because these languages had borrowed words and constructions from each other over the course of centuries. But with languages as far apart geographically as the Mediterranean and India, borrowing over centuries of close contact was ruled out as an explanation. The idea that languages could have similarities because they had "sprung from some common source" was novel. Jones's much-quoted statement is said to have begun the science of historical linguistics. These discoveries were equivalent to DNA profiling of our age; they so enthralled Britain that at Jones's public lectures in later life, over a thousand people would show up. What was discovered, and over the next century more fully pieced together, was the prehistory of a set of languages now distributed from Europe to North India. The original language, which Jones speculated might once have existed and may no longer exist, is now called Proto-Indo-European.

The terms in figure 2.1 illustrate the kinds of similarities that needed to be accounted for. Two words with the same meaning in two different but related languages that are later forms of a common earlier form are called "cognates." The word for "two" in Greek and Latin was the same: *duo*. The Sanskrit form was *dva*. Over time the /v/ of Sanskrit had become /u/ in Greek and Latin. Aside from this somewhat minor, but recognizable, sound shift, the three

English	Sanskrit	Greek	Latin	Old German	Japanese
one	ekas	heis	unus	ains	hitotsu
two	dva	duo	duo	twai	futatsu
three	tryas	treis	tres	thrija	mittsu
four	catvaras	tettares	quattuor	fidwor	yottsu
five	panca	pente	quinque	fimf	itsutsu
six	sat	heks	sex	saihs	muttsu
seven	sapta	hepta	septem	sibum	nanatsu
eight	asta	okto	octo	ahtau	yattsu
nine	nava	ennea	novem	niun	kokonotsu
ten	dasa	deka	decem	taihum	to

Figure 2.1 The numbers one through ten in five Indo-European languages and also, for contrast, in Japanese, which is not a member of the Indo-European family.

words were clearly cognates. If English "two" is also a cognate, we have to account for two sound shifts: /v/ to /w/ and /d/ to /t/. Linguists know that these are easy shifts for languages to make. /V/ and /w/ are made in similar ways in the mouth, and easily shift from one to the other. (Try it and see). In North Indian languages today, individual speakers vary in their pronunciation of /v/ on a range from a good, hard /v/ to a soft and liquid /w/. Similarly with /d/ and /t/, both are made with the tongue on the alveolar ridge (just behind the teeth) and by stopping—then freeing—the breath; the only difference between the two is a lack of voicing on the /t/.

In the case of *dva/duo/twai* you can probably figure out which of these languages English is closest to. The "spelling" of the English word preserves a pronunciation now lost. If we were to spell English "two" the way we pronounce it, we would spell it "tu." But once we pronounced it *two*, in that pronunciation the similarity to Old German *twai* is apparent. The /d/ has already become /t/, and the /v/ has become /w/. The vowels, /o/ and /ai/ are different, but vowels are a lot more slippery than consonants.

If you were to work out the relationship of the five languages just on the basis of "two," you would probably come up with:

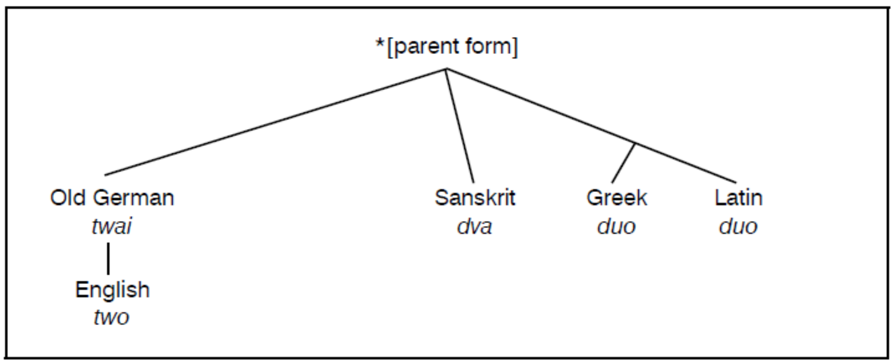


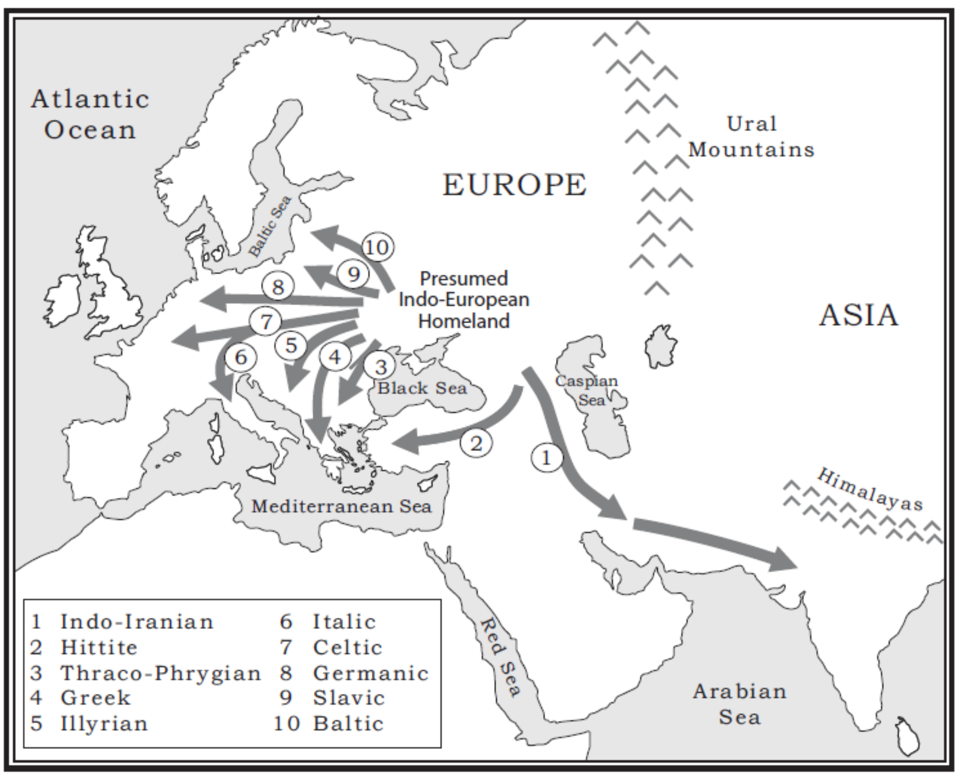
Figure 2.2 Relationship between Sanskrit, Greek, Latin, Old German, and English on the basis of "two."

The tree diagram in figure 2.2 orders the similarities by degree of closeness and by generation. Greek and Latin are identical; between them and Sanskrit there has been one consonant change. Between those three and Old German, there have been two consonant changes. Of course, this is just one word. You would want to take many words to discover repeating patterns or modify your diagram as more evidence seems to require it. Notice, for instance, the /d/ to /t/ shift shows up again in "ten": *dasa/deka/decem* vs. *taihum/ten*. You would probably want to be selective about what words you chose for comparison. The word "motor," for instance, has passed into a large number of the world's languages along with the actual automobile. Languages do "borrow" words extravagantly ("steal" might be a better word; they rarely give them back) along with technological innovations, religious concepts, new foods, etc. Anything new can bring along its label from the original language. These words are called "loanwords." When Christianity moved into Europe, hundreds of religious terms, most of them from Latin, came along with it. The French invaded England in 1066, and over the next three centuries more than 10,000 French words were added to English, often duplicating perfectly good Anglo-Saxon (Germanic) words. Similarly, when Hinduism and later Buddhism spread into Southeast Asia, thousands of Sanskrit words were added to local languages, and Japanese absorbed several thousand Chinese loanwords along with Confucian and Buddhist texts and the Chinese script. When tea was carried westward from China, its Chinese name went along with it. The English and French (*the*) forms are from a Fujian dialect, and Russian (*cha*) and Hindi (*chai*) are from a Cantonese dialect. Now, of course, new words are being created with new technology; words like Internet, software, user-friendly, and cell phone and are not only added to English but are spread widely wherever the new technology travels.

Other types of words rarely change. Pronouns are among the most stable of terms. So are kinship terms, body parts, and words like sun, moon, sky, path, home, name, fire, hot, cold. These are sometimes called "core vocabulary," the oldest, most stable words in any language, and they are used by linguists in comparing languages and finding relationships. These words are most useful when the two "daughter" languages split off long ago. In linguistic terms, a split of 1,000 years is so recent that even untrained observers would not deny they had a common parent. But when the distance to the parent, or "protolanguage," is in the range of several thousand years, it may not be so readily apparent. Most linguists now agree that "Proto-Indo-European" was spoken some 5,000 years ago. This is by far the best-documented language family in the world, partly because so many scholars have worked on it during the last 200 years, and partly because there is documentation of old, intermediate forms in texts written in ancient Sanskrit, Greek, Latin, and Hittite.

In fact, much of the vocabulary of Proto-Indo-European (PIE) has been painstakingly reconstructed, and this reconstructed lexicon holds clues about where PIE was spoken. There were no words for elephant, tiger, banana, rice, or ocean; there *were* words for herd, cow, sheep, pig, goat, dog, house, wolf, bear, goose, duck, bee, salmon, beaver, squirrel, beech, willow, oak, grain, and wheel. It follows that the original Indo-Europeans must have been acquainted with these things, and an effort to map their distribution on a map of Eurasia has led to the hypothesis that the original homeland was the central grassland regions north of the Black and Caspian Seas. It is assumed they were a mounted warrior people who, around 3000 B.C.E., began to move westward into Europe, south into Turkey, and southeast into Iran and India. The Indic, or Indo-Aryan, branch moved into North India around 1700 B.C.E. where they encountered a civilization in decline (the "empty cities" of Harappa and Mohenjo-daro; see chapter 5) and speakers of a wholly different language family, Dravidian. Conquering and intermarrying, they spread across the northern half of the subcontinent, where their language, Sanskrit, began to undergo its inevitable regional transitions into such modern languages as Punjabi, Hindi, Bengali, Marathi, Gujarati, and Nepali.

The language that gave Sir William Jones his great insight, Sanskrit, provides the oldest known samples of an Indo-European language. It continues to be both written and spoken to this day, but it has not been a "mother tongue"—the first language learned by children in a language community—since some time in the first millennium B.C.E. Already by Buddha's time in the sixth century B.C.E. it was archaic, for Buddha insisted on preaching in the vernaculars, the actually spoken languages, and most of the Buddhist texts are written in such a language, Pali. But early Sanskrit and even earlier "Vedic" were preserved by the oral transmission of the most sacred compositions of the Hindus, the Vedas. These four works are regarded by all Hindus as the source of their religion, given directly to humans as "revealed" truth.



Map 2.1 Distribution of Indo-European languages.

East Asian Homelands

The great success of reconstructing the Indo-European language family has naturally led linguists who work in East and Southeast Asia to attempt similar reconstructions there. But in this part of the world the complexities are much greater and the time-depth deeper. There have been far fewer linguists at work on far more languages, and the effort has only begun in the last century. This means that at the present time there is still much uncertainty and disagreement, though some general outlines are beginning to emerge, and we intend to present a picture that is a good deal firmer and clearer than many experts would prefer. This emerging picture has come about as other kinds of scholars, particularly archaeologists, have brought their data to bear on questions also of interest to linguists.

For instance, as soon as languages have been shown by linguists to be members of a language family, we want to know *where* the original mother tongue was spoken, *when* it was spoken, and *why* and *how* the languages began moving to where we now find them. With such questions, obviously, archaeology can be helpful. The archaeological record of excavated sites can be examined for important clues about the people. For instance, were they foragers or cultivators? What is the geographical range within which a typical assemblage of artifacts is found? What is the physical type of the human remains? What seems to have been the population densities? How long were sites inhabited by the same people?

What, for instance, would cause a population to disperse? This is the first stage in the breakup of a language into a number of daughter languages. According to archaeologist Colin Renfrew (1989), there are four principal ways by which populations, together with their languages, spread. First, there were the earliest periods of human dispersal, when people settled in previously uninhabited territories, leaving the African homeland and moving throughout most of Eurasia, Australia, and New Guinea. In some parts of the world, for example the eastern Pacific islands, this has happened recently enough that we have quite a lot of knowledge about the peoples and their languages. In Asia, those earliest *Homo sapiens*, we are certain, must have had some language, if only we could rediscover its traces in modern languages.

This initial migration was followed by the "farming dispersal." The invention of farming in various places caused population growth, and as these agricultural populations expanded at the expense of earlier foragers, they took their languages with them. The full impact of these dispersals is only now coming to be recognized by historical linguists. For instance, Renfrew has argued that the Indo-European dispersal was not by mounted warriors of the steppes but by the earlier agriculturists of the Middle East, very well documented archaeologically, who moved from the Mediterranean north into Europe, displacing earlier foragers (leaving the Basques as a lonely remnant), and also eastward into the Iranian Plateau and India. (Historians of India are not yet accepting this hypothesis, however; if and when they do, it will require radical rethinking of early Indian history.)

The third form of dispersal came about during the global warming several thousand years ago when regions north of the 54th parallel became liveable and pioneers settled in Siberia and crossed the Bering Straits into the circumpolar region where they became the Eskimo-Aleut.

Finally, the fourth form of migration, occurring in later historic times as societies became more complex, is "elite dominance," when incoming groups, better organized and better armed, conquer indigenous groups and imposed their languages on them. The Altaic languages (Mongolian, Manchurian, Turkish, Korean, and possibly Japanese) originated in Central Asia and were carried westward and eastward by well-documented conquests (though Kublai Khan's invasion of China and the dynasty he established, the Yuan, left scarcely a trace on the Chinese language). Thai and Lao spread into Southeast Asia, shoving aside Khmer. The spread of English accompanying the era of colonialism would be a more recent example.

Just as archaeologists know that the Neolithic is invariably accompanied by population growth, linguists know that a region with the greatest diversity of languages is likely to be the oldest area of settlement and possibly the original home of the speakers of a language family. For an easy example that this might be so, think of the diversity of English dialects in the British Isles (in as small an area as London, "Queen's English" and Cockney are extraordinarily different) or the dialect diversity of American English on the East Coast as opposed to the homogeneity of the more recently settled Midwest and West. The same process, allowed to develop over, say, 7,000 years, will turn dialects into languages and then into language families. If a region is an ancient center of agriculture it is more likely to have resisted intrusion from people speaking other languages, and its own language(s) will go on modifying, diversifying, and transforming themselves into a family of related languages. Then, of course, if people begin to move around, at first because of the population explosion associated with agriculture, and later because of military adventuring by conquering peoples, the picture becomes quite complex.

Let us make a simple observation from a linguistic map of today: the greatest linguistic diversity in East and Southeast Asia can be found in a circle with a radius of a thousand miles with its center on Canton. This circle will include the island of Taiwan, the whole eastern region south of the Yangzi, as far inland as Kunming, and the northern hills of Burma and Laos.

This is, first of all, the area where China's many "dialects" have their greatest diversity. Besides Chinese, Southern China also contains four additional language families: Austroasiatic, Tai-Kadai, Miao-Yao, and Tibeto-Burman. And just across the Formosa Straits on Taiwan is a member of the Austronesian family and the probable original homeland of a group of languages now spread from Madagascar to Hawaii.

The later out-of-Africa dispersal of *H. sapiens* traversed southern Asia toward southeast Asia around 60,000 years ago. The southernmost of these people crossed Wallace's Line into Australia and New Guinea, remaining foragers in Australia but independently inventing agriculture in New Guinea. Then, quite recently in this grand view of things, some of those who remained in Asia began to cultivate grain. At about 6000 B.C.E. they were cultivating millet in North China, and about 5000 B.C.E. they were cultivating rice over a region stretching from coastal South China through the hills of Burma and across into India as far as Bihar. We assume that these earliest cultivators were the first to experience the population growth that accompanies agriculture, and the first to disperse amongst foraging peoples, displacing and isolating them. We do not, of course, know anything about their languages. But based on the hard work of many linguists, we can perhaps piece together a picture as things stood between the end of the last glaciation, 10,000 years ago, and the period of the first written texts from China, about the fifteenth century B.C.E.

Several things seem certain to be true during this long period. First, as the area being discussed is vast, it is unlikely that peoples living there 10,000 years ago would all have been speaking the same language. However paleolinguists eventually settle the great debate on the origin(s) of human language, by 10,000 years ago there must have been several, and perhaps many, languages in existence in East and Southeast Asia. Second, no one could then have been living in the fully glaciated Himalayas; settlement of the Himalayas by Tibetans and others came much later. Third, the Neolithic "revolution" must have spread slowly throughout the human communities in this vast area, but as a population acquired the new economy they would quickly have experienced population growth at the expense of others. These newly cultivating populations were Renfrew's "farming dispersal" at the expense of still-foraging populations.

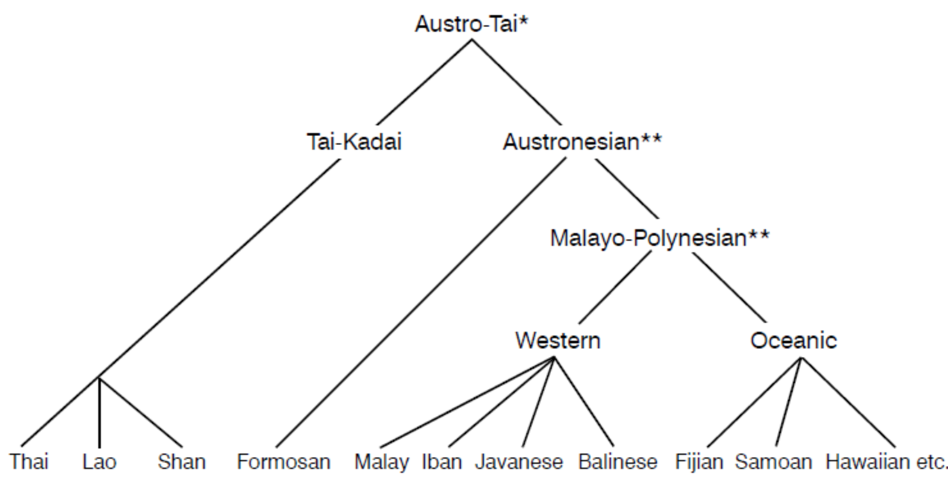
Austroasiatic

There is a family of about 150 languages called Austroasiatic that we assume must be the languages of this first population explosion; it includes Vietnamese and Khmer in Southeast Asia. Over in India it survives among a few hill tribes such as the Munda. But once they must have been widespread throughout the region, including South China; the old name for the Yangzi, *jiang*, is believed to be Austroasiatic, and from their reconstructed vocabulary we find linguistic evidence attesting to their knowledge of rice agriculture. Some cognates of *jiang*, such as *song* (Vietnamese), *krung*, *kroung*, and *karung*, are now general-purpose words for "river."

Austro-Tai

Somewhat later, around 4500 B.C.E., people who spoke another protolanguage called Austro-Tai and also lived in South China began to migrate across the Formosa Strait to Taiwan. Their subsequent migrations are relatively well documented archaeologically, and the linguistic evidence for much of what follows is also strong. Those who remained behind on the mainland of South China stayed put for millennia and in historic times began moving into Southeast Asia as the Thai, the Shan of northern Burma, and the Lao of Laos. Those who went to Taiwan began their linguistic differentiation, and about a thousand years later, possibly around 3500 B.C.E., began moving southward into the Philippines. Here Malayo-Polynesian began to develop. Eventually a western branch moved into the Indonesian Islands of Borneo, Sulawesi, Sumatra, Java, and Bali, while an eastern branch headed out into the Pacific where they became the Polynesian peoples (Bellwood 1992).

The reconstructed protolanguages of Austro-Tai and Malayo-Polynesian provide interesting evidence for the changes in the economies of peoples adapting to an equatorial environment. In the Proto-Austro-Tai lexicon are words for a wet-rice adaptation: field, wet field, garden, plow, rice, sugarcane, the betel nut complex, cattle, water buffalo, ax, canoe. A new set of terms appears in Proto-Malayo-Polynesian indicating their new adaptation in the equatorial



*The Austro-Tai lived in South China prior to 4500 B.C.E. where they developed an economy based on rice cultivation. Their reconstructed lexicon contains words indicating a wet-rice adaptation: field, wet field, garden, plow, rice, sugarcane, the betel nut complex, cattle, water buffalo, axe, and canoe. As populations grew, a group of these Austro-Tai speakers migrated across the Formosa Strait to Taiwan. Others stayed behind, and their languages became the Tai group, now reduced to areas of Southern China near the Laos and Vietnamese border. In historic times some of these Tai began moving into Laos, Burma, and Thailand.

**On Taiwan, over the next thousand years new variants of the language emerged. This became the parent language known as Proto-Austronesian.

***Around 3500 B.C.E. some of these Taiwanese set out for Luzon and settled the other Philippine Islands. The Malayo-Polynesian family began to emerge. As they then spread westward into the Indonesian Islands of Borneo, Sulawesi, Sumatra, Java, and Bali, and eastward into the Pacific, the economy of many of the groups again changed to suit equatorial conditions, and this is reflected in a new vocabulary: taro, breadfruit, banana, yam, sago, coconut. When they entered the Pacific, rice disappeared from their economy and from their language.

Figure 2.3 Members of the Austro-Tai language family.

islands: taro, breadfruit, banana, yam, sago, coconut. Rice disappears from the economy of the Malayo-Polynesians when they enter the Pacific.

Sino-Tibetan

This brings us to Sino-Tibetan, which includes Tibeto-Burman and Chinese. Surprisingly, Sino-Tibetan may be the least understood of all these language families. Though there have been scholars aplenty—for centuries—devoted to the study of Chinese, its vast written traditions have absorbed most of their energies. There has been far less interest in spoken forms of the language, and still less interest in establishing links between the elevated language of the Middle Kingdom and the most certainly "lower" forms of speech of Western and Southern Barbarians. As a result, as recently as a few decades ago, many scholars were still uncertain whether Chinese belonged with the Tibeto-Burman group at all. Tibeto-Burman consists of over a hundred poorly documented languages spoken mostly by small tribal groups, although a few of these, notably Tibetan and Burmese, have very important literary traditions. In fact, a seventh-century Tibetan text and a twelfth-century Burmese text have been useful in reconstructing Proto-Tibeto-Burman.

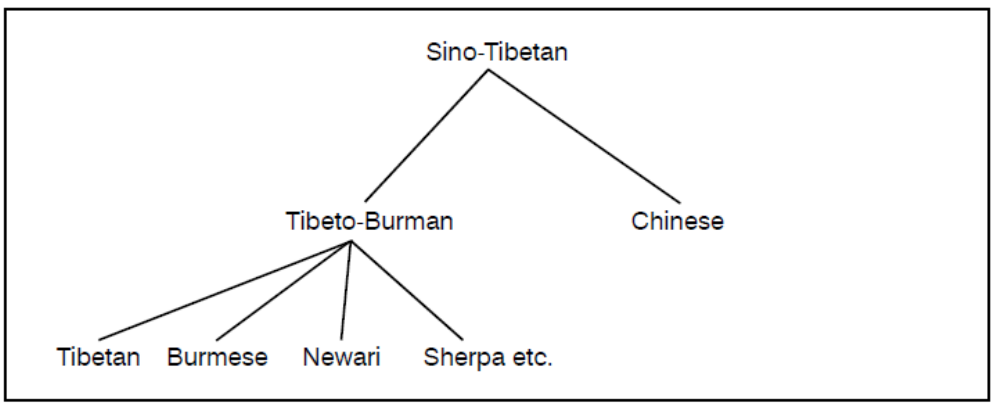


Figure 2.4 Members of the Sino-Tibetan language family.

All the Sino-Tibetan languages are tonal, but so are a number of other nearby language families, like Tai-Kadai and Miao-Yao. Further, Chinese is one of the most purely "isolating" languages in the world; that is, its words, mostly monosyllabic, are very rarely modified with prefixes and suffixes (unlike English, Sanskrit, and Japanese, which are highly inflected languages). This, too, is a Sino-Tibetan characteristic, but Chinese is a more extreme form. However, the word order of the Chinese sentence is unlike the Tibeto-Burman languages: subject-verb-object (SVO) rather than subject-object-verb (SOV). (English also happens to be SVO, a meaningless similarity since there are only six possible variations for all world languages to choose from, and only three are widespread.) That made Chinese the odd one out and led some to doubt the existence of a Sino-Tibetan Family. However, when linguists got down to the exhausting job of comparing the core vocabulary of Chinese with the Tibeto-Burman languages, all doubts about their relationship vanished. Here is a sample:

	Chinese	Tibeto-Burman
sun, day	niet	*niy
name	mieng	*r-miŋ
neck, collar	lieng	*liŋ
tree, wood	sien	*siŋ
year	nien	*s-niŋ
fish	ngio	*ŋya
hair	sam	*tsam

Figure 2.5 Relationship between Chinese and Tibeto-Burman.

Speakers of nontonal languages are usually intimidated by languages with tones and don't quite understand how they work. In Chinese the issue is not really as terrifying as it seems. All languages need many thousands of words in order to talk about the wide ranges of topics humans want to discuss, and these words have to be constructed out of the limited number of sounds in their sound system (i.e., the phonemes). In order to construct enough words, most languages put together multisyllable words. With prefixes and suffixes, these can go to extremes, such as the infamous "antidisestablishmentarianism" in English. Chinese, however, has opted for a lexicon of single-syllable words, but gets quadruple-duty out of them with their four-tone system. Take a simple word like *ba*. English only makes use of this as a quaint Dickensian expletive, as in "Bah humbug!" But in Chinese, meaning differences are conveyed by tones as well as by phonemes. If *ba* is spoken at a high-level pitch, it means "eight." If it is spoken at a high and rising pitch, it means "pull out." If it is spoken at a low falling-then-rising pitch, it means "grasp." If it is spoken with a high pitch that abruptly falls and stops, it means "dam." All of this requires that as a Chinese child learns her language each new word has to be learned with its correct tone, not really such a difficult process. Even so, Chinese appears to have run short on words, as evidenced by the fact that there are a great number of homophones—words with identical pronunciations and tones but different meanings, like English "to," "two," and "too." This has come about because of the simplification of Chinese morphemes over time. Once they had consonant-cluster beginnings (/ts/, /dj/, etc.), and final consonants, but most of these have disappeared, collapsing a set of words into a single consonant-vowel syllable such as those in figure 2.6.

bā	"eight"	bá	"pull out	bǎ	"grasp"	bà	"dam"
xī	"tin"	xí	"mat"	xǐ	"wash"	xì	"opera"
līū	"slide"	liú	"flow"	liǔ	"willow"	liù	"six"

Column one is high level; column two is high and rising; column three is low-falling, then rising; column four is high pitch that falls abruptly and stops.

Figure 2.6 Chinese words indicating the four tones. Column one is high level; column two is high and rising; column three is low falling, then rising; column four is high pitch that falls abruptly and stops.

China's four-tone system is actually a bit on the simple side as tonal languages go in Asia. James Matisoff writes:

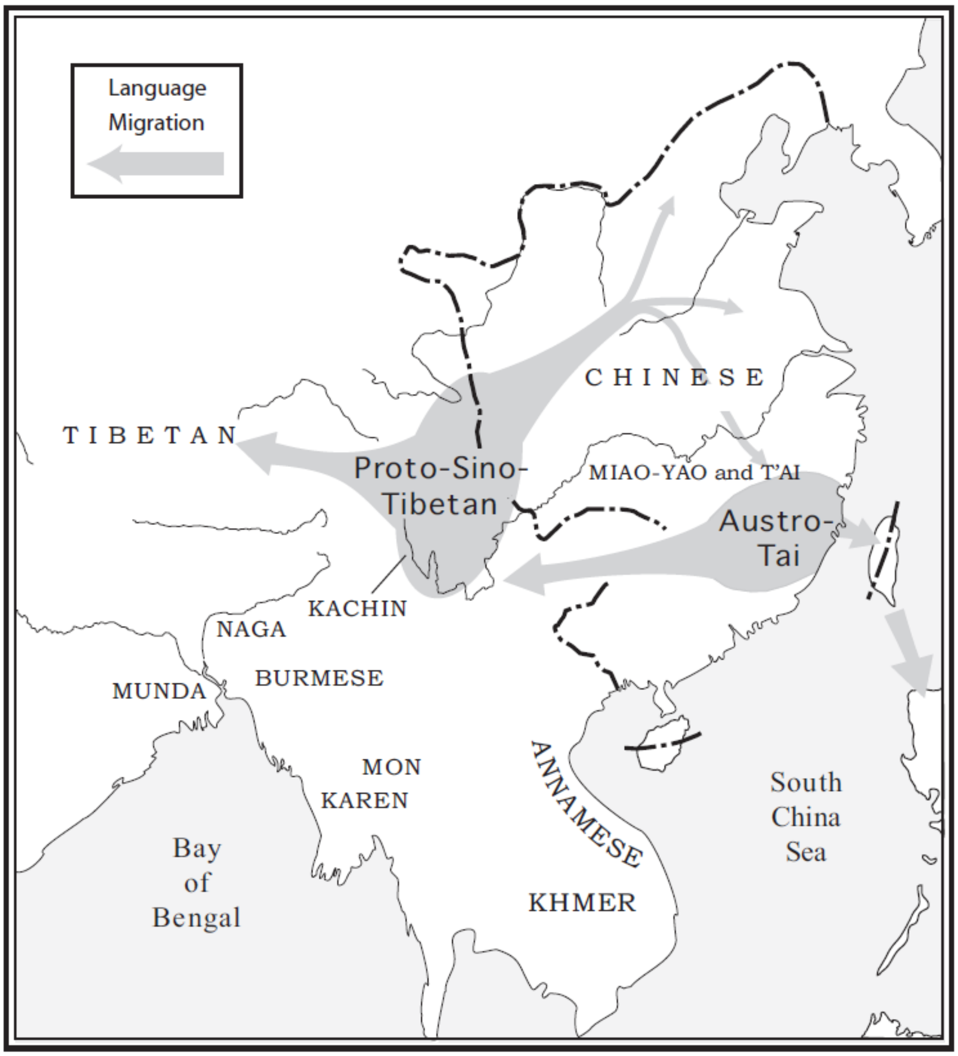
> It is Miao-Yao which is the tonal champion among the tone-prone languages of East Asia. Most dialects of Miao have more than 5 tones, and most dialects of Yao Mien have 6–8, though some have more than 10. The Pu-nu dialect of Yao provides a good example of the tonal virtuosity and "tone-consciousness" of the people of this linguistic area. Pu-nu already has 8 tones of its own in native syllables; but for words which it has borrowed from Chuang and (recently) from Chinese, it has gone to the trouble of introducing three new "foreign" tones, which it keeps distinct from the 8 others! Much can be crammed into the monosyllables of these languages, but there *are* limits! (1983:84)

The above shows how easily languages can borrow tones. It is believed that some Asian languages, such as Vietnamese, which were originally nontonal, became tonal by borrowing tone-systems from their neighbors. Matisoff thinks that Sino-Tibetan was the original "stronghold of tonality" and that the originally atonal Miao-Yao, Tai-Kadai, and Vietnamese acquired their tone systems from the culturally dominant influence of Chinese. The Miao call themselves Hmong, and large numbers of southern Hmong have migrated to the United States from Laos (see chapter 4). Their ancestors may once have been widely distributed throughout North China and perhaps were the original millet cultivators in the Wei Valley. They have a tradition of once inhabiting a snowy climate, and their shamanistic traditions are similar to Siberian shamanism. Thus they have had a long association with tonal Chinese and may have picked up tones from it.

Let us turn to the question of a Sino-Tibetan "homeland." We have seen evidence of a Proto-Austro-Tai–speaking people inhabiting Southeast China who must have been at the coast in order for some of them to spread to Taiwan. Others (the Tai, some of whom later became the Thai, Lao, and Shan) stayed on in Southern China, while some spread to the south and west. It seems likely, then, that speakers of a different language group—Proto-Sino-Tibetan—were settled west of the Proto-Austro-Tai in the foothills of the eastern Himalayas, in

Sichuan and Yunnan Provinces. When the last glaciation receded, some of them must have moved up into the headwaters of the Yangzi, Brahmaputra, Irrawadi, and Mekong around 4000 B.C.E. (Matisoff 1991). From there, one subgroup, the Tibeto-Burmans, followed the Brahmaputra into the Tibetan Plateau, or followed the Irrawadi and Salween south into Burma, where their languages continued to diverge. But another subgroup moved northeast, into the Huanghe and Wei River Valleys. These were the Chinese speakers.

These proto-Chinese speakers were surely not the original inhabitants of this central Huanghe region. The Chinese have no tradition of a migration

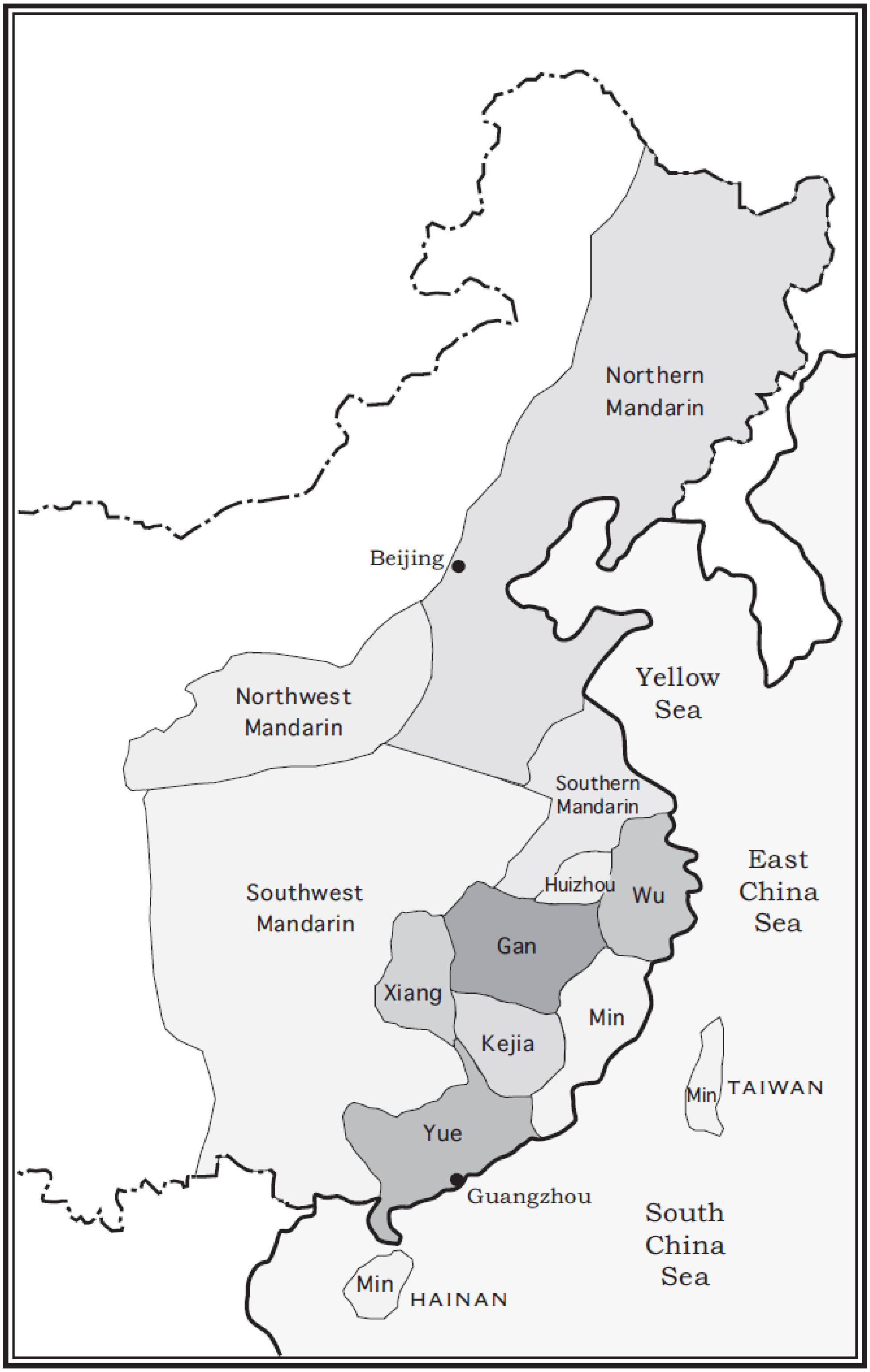


Map 2.2 Hypothetical distribution of Proto-Sino-Tibetan and Proto-Austro-Tai.

from elsewhere, but their consciousness of an "ethnic" difference from other peoples, expressed in early texts, suggests a multilinguistic situation early on. The earliest records speak of a Ti people to the north (possibly Altaic speakers) and a Jung people to the west (possibly Miao) with whom the ruling families sometimes arranged marriages. The original inhabitants of North China may have been the mysterious Miao, "mysterious" because there has been much speculation regarding this last language family from Southern China, Laos, and Thailand to be accounted for. During the first Chinese dynasties, Shang and Zhou, a distinction was made between the aristocrats, designated by the phrase "the hundred surnames," and the peasants, the *min* ("people"). The people who came to know themselves as the "Han" (92 percent of modern China) have probably descended from a mix of early Chinese-speaking elites and the indigenous people with whom they intermarried. In the process they affected each other's languages. The SVO word order of Chinese may have been borrowed from Miao, along with some vocabulary, such as the word for "dog." And Chinese gave Miao tones.

The dialect map for Chinese (see map 2.3) presents us with a puzzle. Notice that the region of greatest diversity is *not* the old center of early Chinese civilization on the Wei-Huanghe Rivers. It is on the South China coast between Canton and Shanghai. The Amoy dialect, just across from Taiwan, is unintelligible to anyone much further than a hundred miles away. If we were to compare it to a dialect map of the United States, this southeast coast would be equivalent to the dialect diversity of our east coast, the oldest settled area, and we might guess that the broad area of Mandarin is, like the Midwest and West, a recently settled region. This guess, however, would be wrong.

The ancestor of all these "dialects" was spoken on the North China plain in the second millennium B.C.E. Since then, according to Jerry Norman (1988), Chinese has evolved under the influence of two competing forces: first, the spread of Chinese to ever more marginal areas, and second, the conscious imposition of a standard dialect—Mandarin—throughout the empire to achieve greater uniformity. Interesting comparisons can be made between Chinese and Roman imperial expansion in the first few centuries B.C.E. and C.E. The language of Rome, Latin, was carried to France and Spain through imperial conquest and colonization. Similarly, early Chinese was carried south and east by the Qin and Han dynasties. For instance, in 221 B.C.E. the First Emperor ordered 500,000 military families to settle in the newly conquered lands of the "Yue people." In the process of colonization in Europe, Latin diverged into Italian, French, Spanish, Rumanian, and Portuguese, while in China, Chinese diverged into Mandarin, Gan, Xiang, Hakka, Wu, Min, and Cantonese. These two families are roughly comparable in age and degree of difference between the daughter languages. However, there the similarity ends. For the Roman Empire never again rose to claim its old territories and assert the unity of its diversifying language(s), to claim them all as mere dialects of Latin; whereas that is exactly what China did. There is one language, Chinese, and its "dialects" are called



Map 2.3 Dialect map of China.

fangyan, “regional speech.” Chinese speak of “picking up” one of the other regional speeches but do not think of themselves as “learning another language” as an Italian learning French would say. Nor has there been a time in Chinese history, even during periods of disunity, when a region asserted its independence by the sort of “linguistic nationalism” that divides India, another place where the Latin model, not the Chinese model, has been followed.

The diversity we see in South China on the dialect map has therefore been growing for the last 2,000 years. At the same time, the dialect of the Beijing area, Mandarin, has dominated since the fifteenth century. Called *guanhua*, the “language of the officials,” it is spoken everywhere north of the Yangzi. It even has its own dialects. Thus, it is repeating the early history of Chinese, moving southward, diversifying, and absorbing earlier dialects in its wake (Norman 1988).

Texts

“You Are Hurting My Language”

The linguist J. L. Becker describes his first lesson in writing Burmese at the beginning of three years of study with a kindly old teacher, U San Htwe:

> As I had been taught to do, I would ask him words for things and then write them down. He watched me writing for a while and then said, “That’s not how you write it,” and he wrote the word in Burmese script. For the word evoked by English “speak,” I wrote / pyɔ / and he wrote ပြော. I insisted it made no difference. He insisted it did and told me I was hurting his language. And so I began, somewhat reluctantly, to learn to write Burmese: /p—/ was a central ပ, and /-y-/ wrapped around the ပ to make ပြ and the vowel / ေ—ာ / fit before and after it: ပြော.
>
> This difference in medial representation made a great difference . . . I could not segment the Burmese syllable into a linear sequence. . . . But segmentation into linear sequence is a prerequisite for doing linguistics as most of us have been taught it. . . . To write my kind of grammar I had to violate his writing.
>
> At first it seemed to me a small price to pay, to phonemicize his language. But over the years—particularly twenty years later, in Java and Bali—I learned how that kind of written figure (a center and marks above, below, before, and after it; the figure of the Burmese and Javanese and Balinese syllable) was for many Southeast Asians a mnemonic frame: everything in the encyclopedic repertoire of terms was ordered that way: directions (the compass rose), diseases, gods, colors, social roles, foods—everything. It was the natural shape of remembered knowledge, a basic icon. (Becker 1995:195)

This respect for writing, even to the single syllable, is pervasive throughout Asia. The mystery of capturing a sound in an inscribed symbol is thought to be divine. The name for the script of Sanskrit and the North Indian vernaculars, *devanagari*, suggests the writing of the gods (Horton 1992), and the earliest form, *brahmi*, meant "of Brahma," God. In early Japan, when writing was first borrowed from China, "its very words were pregnant with spiritual power, called 'word-mana' (*kotodama*)" (Horton 1992). When early writers began writing Japanese rather than just reading and copying Chinese, it was to write Japanese poetry once chanted orally and infused with *kotodama*. And about China, Chiang Yee writes:

> In every district of a Chinese city, and even in the smallest village, there is a little pagoda built for the burning of waste paper bearing writing. This we call *Hsi-Tzu-T'a*—Pagoda of Compassionating the Characters. For we respect characters so highly that we cannot bear them to be trampled under foot or thrown away into some distasteful place. (Yee 1973)

Box 2.1 "Written Symbol" Is Its Name

Greatly may I be forgiven for my intention to call forth a story.
And where dwells the story?
There is a god unsupported by the divine mother earth,
Unsheltered by the sky,
Unilluminated by the sun, moon, stars, or constellations.
Yes, Lord, you dwell in the void, and are situated thus:
You reside in a golden jewel,
Regaled on a golden palanquin,
Umbrellaed by a floating lotus.
There approached in audience by all the gods of the cardinal directions . . .
There, there are the young palm leaves, the one *lontar,*
Which, when taken and split apart, carefully measured are the lengths and widths.
It is this which is brought to life with *hasta, gangga, uwira, tanu*.
And what are the things so named?
Hasta means "hand"
Gangga means "water"
Uwira means "writing instrument"
Taru means "ink."
What is that which is called "ink"?
That is the name for
And none other than
The smoke of the oil lamp,
Collected on the bark of the kepuh-tree,
On a base of copper leaf.
It is these things which are gathered together
And given shape on leaf.
"Written symbol" is its name,
Of one substance and different soundings.

Source: Translation from a Balinese text by Mary Zurbuchen, quoted in A. L. Becker, *Beyond Translation*. Ann Arbor: University of Michigan Press, 1995.

The sacredness of the captured word probably follows from the nature of what was written down in the earliest texts. Although some have argued that writing began in the Middle East as a way for merchants to keep track of their stock and sales, there is no evidence for a similar mundane origin of writing in Asia. Rather, writing began in China 3,000 years ago to communicate with the ancestors; in India it was first used to declare the Buddhist message throughout Emperor Ashoka's realm and then, finally, to write down the sacred Sanskrit hymns of the Vedas that had been chanted orally for a millennium. When writing came to Southeast Asia, it came embodying Hindu and Buddhist teachings. How could the written word not be sacred, when all that was captured in it was?

The Search for Sacred Texts

The scripts with which to record sacred words were invented only twice in the whole of Asia, and all other scripts are adaptations of those originals. But before we turn to that story, there is the story of the texts to be told. For sacred texts spread through Asia in many directions and by many means. Priests carried them to serve kings and religious communities far from home. Pilgrims set out on long journeys seeking authoritative texts from their source. Adventurers sought hidden libraries and plundered them. Poor or unscrupulous dealers trafficked in stolen texts and occasionally created ingenious forgeries. Texts were translated from one language to the next, and the next.

In China, the process of getting Sanskrit texts translated into Chinese was tedious. Few Chinese ever learned Sanskrit, and finding ways to convey multisyllabic Sanskrit words into monosyllabic Chinese was not easy. For instance, the Sanskrit word, *nirvana*, was translated *nie-pan*, meaning "opaque place of retirement." And in the early years the "texts" were often oral, and required "translation teams" of four persons: one to recite the oral Sanskrit, one to write the Sanskrit down (usually two monks from India), one to orally translate the Sanskrit into spoken Chinese, and one to record the oral Chinese into written Chinese (Tanabe 1988). You can imagine how cumbersome this process was. It also shows how closely linked were orality, text-reading, and text-writing. Studies of the "performance of texts" and the "ethnography of reading" (Boyarin 1993; Engelke 2004; Street 2009) show that the silent, lonely room in the mind where a reader is alone with his private decoding of the text and his private "thoughts" is a rarer and more recent phenomenon than we have supposed.

One of the most important of the early text gatherers was a Chinese pilgrim named Xuanzang, who grew up in the chaos following the fall of the Sui dynasty in 618, when the only books being studied, he complained, were military texts. He studied under two great Buddhist monks at Chang'an, but finding the Buddhist texts few, incomplete, and conflicting (probably because of the translation method described above), he resolved to go to India to acquire the originals. He was only 28 years old when he left in 630 C.E., and his lengthy record of this journey is one of the most important sources we have on early seventh-century India. (Xuanzang was not the first Chinese pilgrim to go to

India in search of texts; 200 years earlier, Faxian made a similar journey, also to collect texts and visit the sites associated with the Buddha, and he left an important record of this journey.) Crossing through the thirsty wastes of the great central deserts, endangered by bandits but also aided by devout rulers, Xuanzang finally reached India. There he visited all the major Buddhist sites and spent five years at Nalanda learning Sanskrit and studying the Buddhist texts.

In India, Xuanzang found the Indian method of bookmaking remarkable. Indians wrote on palm leaves and strung the leaves together on a string. Two types of palms, talipat and palmyra, were suitable (in Southeast Asia, the lontar palm is also used). Each leaf had to be cut from its central spine, trimmed to size, boiled in milk or water several times, and finally rubbed smooth with a cowry shell or stone. They could be written on with ink or incised with a stylus, and the type of writing implement affected the emergence of scripts, as we shall see. The disadvantage of palm leaves is their vulnerability to humidity and insects, which is why very few palm-leaf manuscripts older than the sixteenth century exist. They require tremendous care; they are kept wrapped in cloth when not in use and have to be recopied frequently. They have survived better in dry climates outside India and Southeast Asia. In 608, 22 years before Xuanzang left for India, a Japanese envoy brought an Indian palmyra leaf text that he had acquired from Sui dynasty (China) to Japan, and it is still there, the oldest in existence (Tanabe 1988). The fragility of texts probably contributed to the Bud-

Indian palm leaf book in Malayalam script, similar to those seen by Xuanzang in the seventh century C.E.

dhist notion of the fragility of dharma, in continuous decay since the time of Buddha's existence on earth and constantly requiring protection and renewal.

Xuanzang was used to texts written on silk that could be hung or rolled up in a scroll to be stored compactly, although already by his time paper was in frequent use. Because silk was costly, the Chinese invented paper as an alternative as early as 105 C.E., at first by recycling old silk rags and later cheaper linen. The rags were first allowed to rot; then they were cleaned, bleached, pulped, mixed with wheat flour, and finally spread on a frame to dry in sheets. These sheets were then glued together in great lengths, typically 20 feet long, which could be rolled up like silk scrolls for storage (Tanabe 1988).

Xuanzang was a member of the intellectual elite of his time, in search of philosophical-religious texts, but these texts and the arguments they contained were embedded in a worldview where everything associated with the Buddha—himself no longer in existence—radiated power, or spiritual merit. As a devotee, Xuanzang describes being overcome with emotion at sites where Buddha once taught or where relics of the Buddha's body were enshrined in a *stupa*. He later recalled the power of these places: "A mysterious sense of awe surrounds the spot; many miracles occur. Sometimes heavenly odors are perceived" (Hopkirk 1980). When finally, 15 years later, in 645 he returned to the Monastery of Extensive Happiness in Chang'an, he deposited six statues of Buddha in various acts, 657 texts carried on 20 horses, and 150 particles of the Buddha's flesh. We should not think of these as eccentric or grisly souvenirs, but as dharma itself transported to China. The objects were at once symbols, carriers, and extensions of dharma. The texts, the Buddha relics, and the images shared in the power of Buddha himself.

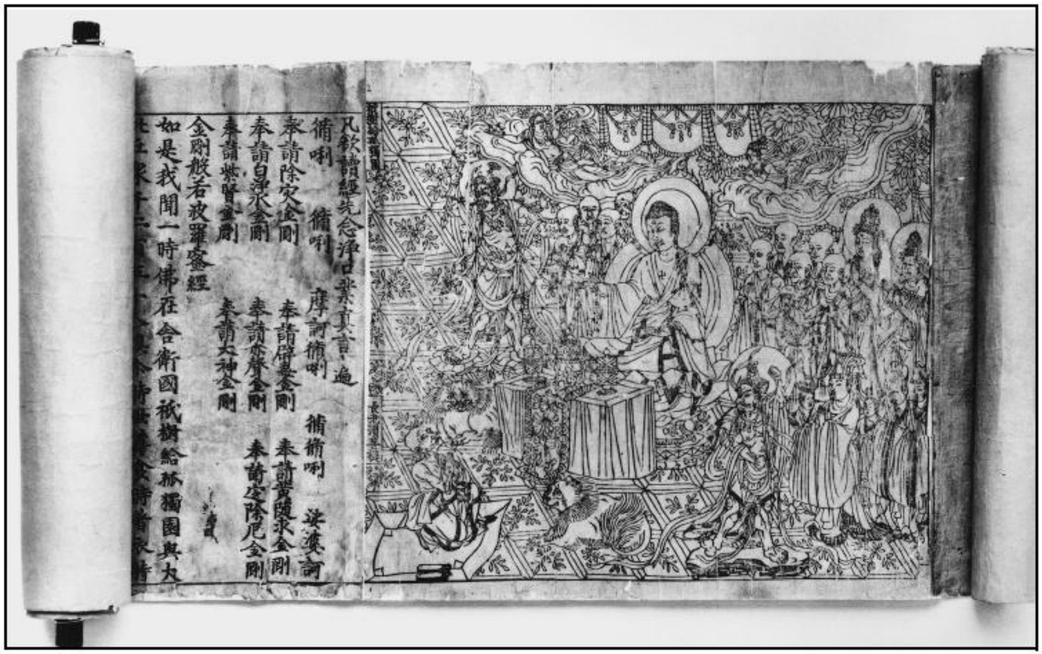
This began a whole new era of translating and then of "sutra copying," both in China and in Japan. There was first of all the task of translating the new works brought back by Xuanzang. In catalogs of Buddhist texts made two centuries later, the Chinese Buddhist canon had grown to 1,076 and then to 1,258 works. At first the concern was just to get good Chinese translations in an efficient manner, but later it became an act of religious merit-making to have them copied. The wealthy employed monk-scribes, or if they were talented, copied the sutras themselves. Emperors commissioned copies of the entire canon, and for this, large official bureaus were set up and permanently staffed. Finally, attention turned to making them beautiful. The first sheets of scrolls were often decorated with scenes from the sutra that followed, and sometimes the paper was dyed a deep indigo and the sutra written in gold characters. Many of these have survived in Japan, with a postscript noting that the sutra had been commissioned by Empress Shotoku or Lady Fujiwara or Emperor Shomu to transfer merit to specific ancestors.

One of the places Xuanzang stopped for a little while on the outward and return trips was Dunhuang, at that time China's western outpost, the "Jade Gate" through which caravans left for the perils of the Taklamakan and the western world. Here there was a great Buddhist monastery carved into a mile of stony cliffs known as the "Caves of the Thousand Buddhas." Four hundred and sixty-nine of more than a thousand caves still exist. For 1,500 years monks and devout laymen filled these grottoes with paintings and sculptures and earned merit by commissioning copies of sacred texts.

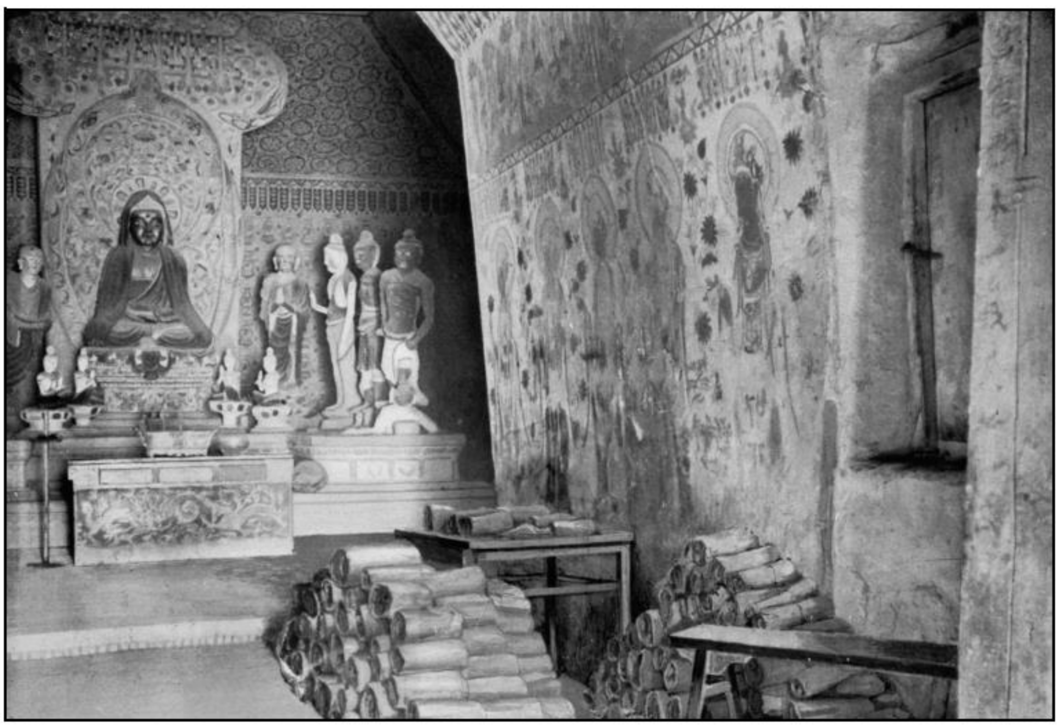
Dunhuang continued in use, though much reduced in importance, into the twentieth century, unlike many similar sites more in the reach of the desert. In one of the first years of the twentieth century, a Daoist priest named Wang Yuanlu made a stupendous discovery. While carrying out an ambitious but underfunded and poorly conceptualized revamping of some of the caves, he discovered a concealed room that had been bricked over around 1000 C.E. Yanking out some of the bricks, he was astonished to find thousands upon thousands of manuscripts, a whole library, in a dozen languages or more. Wang was no scholar, and actually reading these scrolls was not high priority for him. But he reverenced them because they contained the teachings of Lord Buddha and the labor of unknown thousands of monks who had lived and studied at Dunhuang since the fourth century, and he was willing to part with a few of them—there were so many!—in order to fund the repainting of frescoes and other repairs he was trying to carry out. He notified the local authorities of his find, but they did not seem to have any clear idea about what should be done next—notifying scholars in Beijing does not seem to have occurred to them—so they told Wang to leave the manuscripts where they were.

Not long after, in 1907, the British explorer and archaeologist Sir Aurel Stein arrived in Dunhuang by camel in the middle of a freezing sandstorm. He was in a fine mood because he had just discovered the long-lost westward extension of the Great Wall mentioned in Chinese annals. What happened next still infuriates the Chinese, and appears to have shamed even Britain, for in the Central Asian gallery of the British Museum, where many of the manuscripts ended up, it is difficult to find even a trace of Sir Aurel Stein. Stein got wind of Wang's finds, and together with his Chinese assistant stayed around for several weeks, trying to find a way to weasel some of the manuscripts from Wang. Stein, it must be said, had as much respect for ancient manuscripts as anyone, though for scholarly reasons rather than for religious ones. He truly believed that to acquire the manuscripts and take them to India, and then to London, was to rescue them from oblivion and possible destruction.[1] Moreover, he considered Xuanzang his patron saint, because like the ancient Chinese pilgrim, Stein, too, crisscrossed the dangerous deserts of Central Asia looking for traces of the Buddha and his followers. This was a happy coincidence, because a local artist had just finished murals commissioned by Wang depicting scenes in the life of Xuanzang. So imagine the impact it had when the very first manuscript Wang drew at random from the pile to allow Stein to examine carried a final colophon that the translation had been done by Xuanzang himself from originals he had personally carried from India.

Stein describes the scene in the little room: "Heaped up in layers, but without any order, there appeared in the dim light of the priest's little lamp a solid



The Diamond Sutra, the oldest printed book in the world. Chinese characters on a rolled-up scroll, copied by Wong Jei in 858 C.E.



In this photo, taken at Dunhuang in 1907 by Aurel Stein, the secret door to the hidden chamber can be seen on the right. Found a short while earlier by a Taoist priest, the chamber, containing thousands of manuscripts, had been walled up 900 years earlier. Some of the manuscripts are piled outside the door for inspection by Stein. He managed to talk the elderly caretaker out of 13,000 manuscripts, most of them now in London or India.

mass of manuscript bundles rising to a height of nearly ten feet, and filling, as subsequent measurement showed, close on 500 cubic feet." He managed to talk the dedicated priest out of 13,000 manuscripts, 7,000 in Chinese, the rest in other languages. Because Stein could not read Chinese, his selection was on the basis of age and condition more than on the importance of the text. Had he known that a thousand of them were copies of the Lotus Sutra, he might have been more selective. The most famous of all was a copy of the Diamond Sutra. Although there are many manuscripts of the Diamond Sutra in existence, this work is famous because it was block-printed on seven pieces of paper joined together with a block-print illustration on the first sheet and bears a date, 868, which makes it the oldest printed book in existence. It can now be seen in the British Museum near the Gutenberg Bible. A Chinese scholar in a National Library of China publication describes the Diamond Sutra and adds: "This famous scroll was stolen over fifty years ago by the Englishman Ssu-t'an-yin [Stein] which causes people to gnash their teeth in bitter hatred" (Hopkirk 1980).

Scripts

South Asian Scripts

The two major families of scripts that dominate in Asia are based on principles different from the Roman alphabet that we use. The alphabet used for writing English is phonetic; theoretically, there is a graph for each sound, both consonants and vowels. (In what follows we will use the term "graph" in place of such diverse, and potentially confusable, terms as letter, character, sign, symbol, etc.) In reality, the correspondence between the actual sounds of spoken English and the 26 graphs available for writing it is not good. What sound does /c/ correspond to? Is it /k/ as in "picnic?" Or is it /s/ as in "implicit"? We could really use a separate graph for /ch/ as in "church," and while we are at it, we could add a single graph for the /sh/ as in "shame," something to stand for the soft /g/ in "corsage" (for people who don't already pronounce it with a /j/), and for the /ng/ in "singing." Perhaps we are most in need of graphs for vowels, such as the different /a/'s in "father," "famous," and "fashion." However, imperfect though it is, the logic of our system is one-sound-one-graph, and vowels count the same as consonants.

We have already seen a different system at work in Burma. That system focuses on the syllable, and the core of the syllable is its consonant. The sense of speech is different in these languages. Speech is made up of so many consonants modified by vowels. In writing speech, you write consonants, then put marks above, below, behind, and in front of them to indicate vowels. So in the Tamil script of South India, /pa/ is ப, /pā/ is written பா, /pi/ is பி, /pu/ is பு, /po/ is பொ, /pai/ is பை, and /pau/ is பௌ. Prior to the advent of computer software with their hundreds of specialized fonts, typewriters for these languages allowed you to make a consonant sign, then instead of moving ahead automatically, it waited while you decided where the vowel marks go.

This column was erected around 250 B.C.E. by Emperor Ashoka, declaring his commitment to Buddhist values. It contains the earliest examples of the *brahmi* script, a forerunner of the modern *devanagiri*.

The syllabic systems of Tamil and Burmese came originally from India. The earliest known Indian script is found on 40-foot-high columns set up by King Ashoka around 250 B.C.E. to mark the boundaries of his empire and to proclaim Buddhism as the law of the land. This script, now known as *Ashokan brahmi*, was not deciphered until the nineteenth century by James Prinsep who worked for the East India Company. A sure sign of its antiquity was the fact that even the Brahmans could not read it. When the first Englishmen saw it, they guessed it might be Greek, but how Greek came to be written on columns in Delhi, Allahabad, and Bihar was anyone's guess. What is usually needed in deciphering ancient scripts is a key text such as the famous Rosetta stone where the same inscription is recorded in a known script and in the unknown one. It turned out that Ashokan Brahmi was not Greek, but an ancestor of Devanagari in which Sanskrit is written, and there was an intermediate script, now known as *Gupta brahmi*, in use during the Gupta Empire in 319–550 C.E. Once Gupta Brahmi was translated, scholars could work backwards to Ashokan Brahmi. The key for Prinsep was a number of stone columns that circle an ancient Buddhist *stupa* at Sanchi in Central India. He wrote to a friend: "The Sanchi inscriptions have enlightened me. Each line is engraved on a separate pillar or railing. Then, thought I, they must be the gifts of private individuals where names will be recorded. All end in *danam* [i.e., the Brahmi graphs for *danam*]—that must mean 'gift' or 'given'" (Keay 1988:52). So *danam* gave him the *d, n*, and *m*, and from there he was able to decode the rest.

Below is a line of Brahmi script. You will notice that there is one character per syllable, with a special graph for a vowel that begins a word: u-pa-sa-ka, etc. Non-initial vowels are indicated by modifying the consonants with little added marks.

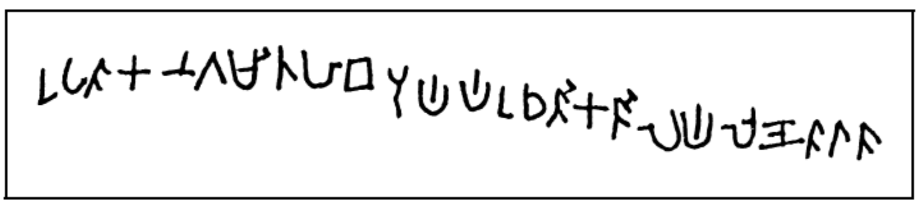


Figure 2.7 Brahmi inscription reading: "The Cave of the female lay-devotee, Shila, wife of the lay-devotee Nagamitta, [is given] to the Sangha."

For instance, the third graph, 𑀰, is /sh/. Without any modification, the vowel associated with *sh* is assumed to be *a*; this is called the "inherent *a*." Thirteen graphs later, it reappears with a little 𑀺 on the top, making it */shi/*: 𑀰𑀺. The vowel /i/ appears with */mi/* 𑀫𑀺 and */ri/* 𑀭𑀺. The vowel */e/* is seen in the second */l/*: */la/* 𑀮 and */le/* 𑀮𑁂, and in */ne/*: 𑀦𑁂. With just this much knowledge, you could predict other Brahmi syllables.

It is unknown where Brahmi originated. Some have supposed that the still-undeciphered script of Indus Valley (see chapter 5) may be its ancestor, but most scholars doubt that. It has important resemblances to Semitic scripts of the Middle East, especially in its emphasis on consonants. Semitic scripts in their earliest forms did not show vowels at all; you had to guess that *k-t-b* meant *kitab*, "book." Later, little marks were placed above, below, and beside the consonants to indicate vowels, and this is the form in which it must have been brought to India. But there is no historical trace of its route or date or means of arrival in India. At any rate, this is the script used by Ashoka, and in later forms of it, was the script of the earliest Hindu and Buddhist texts. All major scripts of South and Southeast Asia descended from Brahmi, but of course they had to be adapted to the very different languages that they had to serve. In Southeast Asia, where the materials for writing were until very recently the palm leaf and iron stylus, the graphs evolved in rounded forms that would not damage the leaves. In North India, where pen and ink were used, the distinctive "clothes hanging from a clothesline" style developed, where the line over the top connects graphs for individual words. Such a line, made with a stylus, could damage the palm leaf. Like Brahmi, each consonant graph has either an inherent *a*, or an attached sign to indicate the vowel:

क	कि	की	के	कु	कू	को	कौ
ka	ki	kī	ke	ku	kū	ko	kau

Figure 2.8 Devanagari script. Sample shows /k/ with vowel markings.

Written Chinese

The word for "civilization" in Chinese is *wenming*. In written form it is composed of two graphs: the graph for *wen* ("writing") and the graph for *ming* ("understanding"). To be civilized has always been a marked feature of Chinese self-identity; to be civilized was to understand writing. The Chinese have had a fully developed script since at least the fourteenth century B.C.E., and emergent forms can be traced back 6,000 years. The most remarkable fact is that this script has been in continuous use ever since, and the script used in the twentieth century is a direct descendant of the earliest forms.

The discovery of the earliest forms of Chinese writing came about much the same way as the discovery of Peking Man fossils: Ancient specimens inscribed on bones and tortoise carapaces were classified as "dragon bones" and ground up as a cure for fever (Chou 1979). In the year 1899 a Peking scholar named Wang Yirong was taking doses of "dragon bones" for his malaria. When he ran short, he asked his house guest, who was also a scholar, to go to the apothecary for a refill. As this man watched the "dragon bones" being freshly ground, he noticed strange markings on them and had a closer look. He thought the markings looked like old forms of Chinese writing, but they were not immediately readable. Wang and his friend bought all the remaining fragments of bone and then asked about the source. Over the next few years this amazing discovery led Chinese scholars to rediscover the ancient history of their own written language. In excavations carried out between 1928 and 1937 at five sites, all around what proved to be the ancient capital of Shang dynasty, Anyang, over 20,000 specimens were found, and eventually the number grew to 100,000.

The ancient Chinese communicated with their ancestors and occasionally with a shadowy deity named Shangdi ("Emperor of Heaven"), by writing questions directed to them on the shoulder blades of oxen or on the carapaces of turtles. These have come to be called "oracle bones." The questions were written twice, once in the affirmative and once in the negative: "Will Lady Hao be in good health after she gives birth to her baby?" "Will Lady Hao not be in good health after she gives birth to her baby?" These were then heated by the method of holding a red-hot iron to the backside of an inscription until the bone cracked. The diviner looked for—and tried to produce—a crack shape that would resemble the Chinese character *pu*, which meant an affirmative answer. Thus, the question was conveyed and the ancestor replied.

Knowing how the Roman alphabet evolved is not particularly helpful in understanding any of its written languages, but for an English speaker trying to understand Chinese writing, it is helpful to know how the graphs came to be formed in earliest times.

The Chinese graph is "logographic"; that is, each graph represents a word, and as we have already seen, in Chinese each word is generally a single syllable.[2] There are very few words like the English word "river" or "elephant" where meaning is carried across two or three syllables. There are two types of logographs in Chinese, simple and complex. Box 2.2 shows some simple ones. The first column consists of the ancient form, the second column the modern form, and the third column the pronunciation in Mandarin. A large number of the earliest simple graphs were fairly obvious pictures of the things they referred to: the sun, rain falling from a cloud, a man walking, a woman kneeling, a child with a head large in proportion to its body, a man in a cell, a kneeling man with his hands raised in surrender, etc. Graphs for less easily drawn words could be made by combining them: "love" is the graph for mother plus the graph for child. "Prisoner" is a man in a cell; "east" is the sun behind a tree; "quarrel" is two women side by side; "flute" is a mouth with a reed, and so on. There are a thousand or so of these simple graphs.

A writing system cannot get very far with pictures alone. To make up more words, scribes began adapting words that were picturable to make words that sounded similar (or at least they did in the language as it was spoken at that

Box 2.2 Simple Chinese Logographs in Ancient and Modern Form

日	ɽɯ̀	sun/day
雨	y̌	rain
水	ʂuĕi	river/water
魚	ý	fish
人	ɽə́n	man (i.e., human being)
女	ny̌	woman
口	k^{h}ŏu	mouth
目	mù	eye
眉	méi	eyebrow
木	mù	tree
其	cī	winnowing-basket
子	çɯ̆	child
兼	ciēn	to have two at once (hand-holding two arrows)
若	ɽuò	yield, conform
東	tūŋ	east (sun behind tree)
囚	ciōu	prisoner (man in cell)
好	xàu	love (woman and child)
奻	nuàn	quarrel (two women)
言	jén	flute
辟	pì	prince

Source: Sampson, 1985, p. 24

time). Let's look at how this worked using English rather than Chinese. In English the word "sun" (picturable) sounds like the word "son." You could use the sun graph for both, and let the correct meaning be deduced from context. You could go on with this process:

= sun
= son
= soon
= sung
= sin
= shun

Graphs like this were roughly *phonetic*; a single graph that once stood specifically for a single word came to stand for a whole set of similar-sounding words. But you can imagine that if the system kept on like this, it would eventually get tremendously confusing. So ancient Chinese scribes developed another device for sorting them out. Another set of elements, known as "significs," were developed to combine with phonetics to give the reader a clue as to which meaning was intended. For instance, the graph for "man" could be combined with the graph for "sun" to indicate "son"—i.e., "the man-type of sun." A graph for an open mouth could be combined with "sun" to mean "sung." The mouth and man graphs could also be combined with other elements to create other words. Because of these combinations, Chinese graphs can be learned and remembered more easily than might be imagined, since constituent elements show up again and again.

All these combined elements are written with great elegance in a single imagined square, so they all take up roughly the same amount of space, unlike English words that vary tremendously in the space it takes to write them out. A Chinese graph may have two strokes, such as in the graph for "man," or 30 or more in a very complex graph, but they will both take up the same size square. These are then written in vertical rows beginning on the right hand side. And because of the monosyllabic nature of spoken Chinese, each graph will (usually) be pronounced as a single-syllable word.

One of the virtues of the Chinese system of writing is the very fact that there is no direct connection between the graph and its pronunciation. This means that any given graph can be called by whatever word is used in the various "dialects" of Chinese, so that it does not matter if all Chinese do not speak the same dialect; they all read the same script. It works much like the numerals in Europe, where the numeral 1 is called "one" in English, *un* in French, *eins* in German, *uno* in Spanish. This feature may be the single most important factor in the historic unity of the Chinese cultural world.

In the last several decades, the People's Republic of China has attempted two far-reaching reforms of the written language. The first reform is an attempted simplification of many of the old multistroke graphs. Simplified Chinese characters have become standard throughout mainland China, but most Chinese communities outside of the People's Republic, such as Hong Kong, Taiwan, and overseas Chinese, have all maintained the use of traditional characters. The second reform has to do with the way spoken Chinese is transcribed into the Roman alphabet. The old system, known as the Wade-Giles system, involved many digraphs and apostrophes, and was often quite misleading as to actual pronunciation. Mao Tse-tung, T'ang, *ts'ao-shu*, *jen* (pronounced "run"), and so forth, are examples of Wade-Giles. The more recent system, pinyin, was adopted in 1958 and is the method Beijing has urged all writers and publishers everywhere to adopt. However, it is questionable whether this new system is really less ambiguous. Most non-Chinese speakers find *qin* and *xia* more puzzling than the Wade-Giles *ch'in* and *hsia*. (In this book we are usually following pinyin except in a few well-known Wade-Giles transliterations.)

Korean and Japanese

Both Korea and Japan were deeply influenced by China when, in the first few centuries of the Common Era, they began forming state systems. China was the sun and they were moons reflecting her cultural glory. Each copied or attempted to copy major institutions wholesale. Korea tried to become a *sohwa*, a "small China." Among other things, each borrowed China's script, having none of its own.

This was probably not a wise thing to have done, if looked at on purely linguistic grounds. Chinese writing was wonderfully suited to the grammatically simple, tonal, and monosyllabic nature of the Chinese language, but Korean and Japanese had not even the remotest connection to the Sino-Tibetan family. Korean is one of the Altaic languages (including Mongolian, Tartar, and Turkish), and Japanese may be also, although there is a lot of uncertainty about that right now. At any rate, both languages are polysyllabic, atonal, and highly "inflectional"; that is, verb forms change by adding a variety of different endings to perform various functions in the sentence. This is something written Chinese does not have the capacity to do.

However, Chinese was the only script available during the critical period. And so both Koreans and Japanese struggled for centuries to make Chinese writing work for their very different languages. The result in Japan was "quite an astonishingly complicated method of making language visible" (Sampson 1985).

There was another factor beyond the language difficulties contributing to the complexity of written Japanese. To quote Geoffrey Sampson:

> Japanese society, during much of the period in which the script was developed, was characterized by the existence of an aristocratic class many members of which lacked political power or indeed any serious employment, so that their only role in life was as definers and producers of cultural norms, ways of civilized living. As a natural result, many aspects of Japanese culture, including its writing, were greatly elaborated, made exquisite and intellectually rich rather than straightforwardly functional. (This con-

> trasts with the case of China, which at most periods of its history was a rather down-to-earth, workaday civilization and where the script, for instance, was shaped in the historical period largely by civil servants who had plenty to keep them busy.) (1985:172–173)

This contrast, however, is no longer true. The Japanese writing system, complicated though it is, works fine for one of the most hard-working and technologically advanced societies in the world. Moreover, Japan has one of the highest literacy rates in the world; at 99 percent, it is higher than the US, France, and many other industrialized nations. And dyslexia is practically unknown.

Let's take it one step at a time. First, Japan adopted Chinese writing along with a whole literature—Buddhist and Confucian texts, poetry and prose—in Chinese. Inevitably, an enormous number of Chinese loanwords merged into the Japanese lexicon during those first few centuries (from the seventh century on). Even today, when the Japanese need to coin new words, they turn to Chinese roots as we turn to Greek and Latin ones. The graphs, taken more or less directly from Chinese and still used in that form, are known as *kanji*, from the Chinese *hanzi*, "Han characters."

When Japanese tried to write native words and sentences, the problems were immense. While the writing system was still the plaything of an idle elite, ingenious adaptations were created, which made reading a matter of clever decoding and elegant word play. But eventually a syllabic script emerged. The Koreans had invented a syllable system (not Hangul, an earlier one) in which the first four syllables were *ka-na-ta-ra*, and the Japanese used the first two, *kana*, to refer to their own similar efforts (in the same way "alphabet" comes from the first two letters of the Greek alphabet). Each *kana* graph stands for a consonant + vowel combination since almost all Japanese syllables take that form. There are actually two *kana*s: *hiragana* (plain *kana*) and *katakana* (partial *kana*). Both sets were constructed by simplifying *kanji* graphs, and both are perfect matches for each of the 49 syllables of the Japanese language.

In modern writing, all three can be used together. *Kanji* graphs are used for morphemes like verb roots and nouns. *Hiragana* is used for inflections at the ends of verbs and nouns (like "-ing" and "-ed" in English) and also to spell out grammatical morphemes that are equivalent to English "of," "the," etc. *Katakana* is used for loanwords from foreign languages other than Chinese and to spell out foreign names.

Korea went through much the same trials for centuries but finally, in the fifteenth century, King Sejong (1418–1450) invented a new script. He set up a "Bureau of Standard Sounds" and appointed a group of scholars to create a phonetic (i.e., syllabic) script. By this time, of course, the Koreans were quite familiar with such scripts, originating in India, which were widely used throughout Asia. In two years, a 24-character system was devised that has been called "the world's best alphabet." It took a while for the script to be accepted, because the educated elite viewed it as a trivialization of Chinese. They called it "vernacular writing" (*onmun*), and for a long time it was used only for the most practical clerical functions, a kind of inferior shorthand. But in this century the script was renamed Hangul, "great script," to raise its status, and it is used exclusively in North Korea and widely in South Korea.

ENDNOTES

[1] There was something to this concern. A few years later, 400 White Russian soldiers were interned in the caves. They did so much damage to the great wall paintings that now Stein's photographs are the only remaining record of them.

[2] The view that each graph represents an idea or concept rather than a word is quite misleading, therefore the term "ideograph" is best avoided.

REFERENCES CITED

Becker, A. L. 1995. Biography of a Sentence. In *Beyond Translation; Essays Toward a Modern Philology,* ed. A. L. Becker. Ann Arbor: University of Michigan Press.

Bellwood, Peter. 1992. Southeast Asia Before History. In *The Cambridge History of Southeast Asia*, Vol. 1, ed. Nicholas Tarling. Cambridge: Cambridge University Press.

Boyarin, Jonathan, ed., 1993. *The Ethnography of Reading,* Berkeley: University of California Press.

Chou, Hung-hsiang. 1979. Chinese Oracle Bones. *Scientific American* 240(4):135–149.

Engelke, Matthew. 2004. Text and Performance in an African Church: The Book, "Live and Direct." *American Ethnologist* 31(1): 76–91.

Franklin, Michael J. 1995. *Sir William Jones: Selected Poetical and Prose Works.* Cardiff: University of Wales Press.

Hopkirk, Peter. 1980. *Foreign Devils on the Silk Road: The Search for the Lost Cities and Treasures of Chinese Central Asia*. London: Oxford University Press.

Horton, H. Mark. 1992. Japanese Spirit and Chinese Learning: Scribes and Storytellers in Pre-Modern Japan. In *The Ethnography of Reading*, ed. Jonathan Boyarin. Pp. 156–179. Berkeley: University of California Press.

Keay, John. 1988. *India Discovered.* London: Collins.

Matisoff, James A. 1983. Linguistic Diversity and Language Contact. In *Highlanders of Thailand*, ed. John McKinnon and Wanat Bhruksasri. Pp. 56–86. Oxford: Oxford University Press.

———. 1991. Sino-Tibetan Linguistics. *Annual Reviews in Anthropology* 20:469–504.

Norman, Jerry. 1988. *Chinese.* Cambridge: Cambridge University Press.

Renfrew, Colin. 1989. The Origins of Indo-European Languages. *Scientific American* 261(4): 106–114.

Sampson, Geoffrey. 1985. *Writing Systems.* London: Hutchinson.

Street, Brian, 2009. Ethnography of writing and reading." In *The Cambridge Handbook of Literacy,* ed. David R. Olson and Nancy Torrance. Pp. 329–345. Cambridge: Cambridge University Press.

Tanabe, Willa J. 1988. *Paintings of the Lotus Sutra*. Boston: Weatherhill Press.

Yee, Chiang. 1973. Chinese Calligraphy: An Introduction to Its Aesthetic and Technique. Cambridge: Harvard University Press.