A HISTORY OF THE CHINESE LANGUAGE

HONGYUAN DONG

Contents

Figures

Tables

Preface

Language is one of the fundamental aspects of human society, and language history is a fascinating topic both in itself and in relation to other subjects such as history, literature, archaeology, etc. and therefore studies of the history of language deal with linguistic matters per se and with relevant sociocultural factors. For learners of a modern language, knowledge of the historical development of the sounds, grammar and vocabulary can also be helpful since, among other things, classicisms are quite common in many modern languages. This is especially true with languages like Chinese, where the classical language had been used as the official written language for about two thousand years before modern times. Anyone who is interested in Chinese language and culture and in linguistic history should find this book useful.

Chinese historical linguistics is a difficult subject. There are some very good introductory books available but they are either too advanced for the uninitiated or too sketchy for the motivated and so I want to make this subject more accessible to the general readership. I will not only describe the features of the different stages of the Chinese language, but also guide the readers step by step through many interesting findings.

Readers with no prior knowledge of linguistics or Chinese can make use of the appendices. In Appendix I, I have compiled a list of major chronological divisions in Chinese history. Most of the symbols used in this book are International Phonetic Alphabet (IPA) symbols and there is a short tutorial on phonetics and the IPA in Appendix II. The pronunciation of Modern Standard Chinese is given in *pinyin*, the Romanization system currently used in China, and there is a pronunciation guide to such pinyin symbols in Appendix III. There are also two pinyin-IPA correspondence charts in Appendix IV. A quick reference to these appendices can be made whenever needed.

Pinyin is also used in the transcription of the names of Chinese people and the titles of ancient classics. Some of these have been popularly transcribed in other notations, such as the Wade-Giles system. For example, the name of the ancient philosopher, Laozi, has been variously rendered as Lao-Tsu or Lao Tse, and his book, the *Dao De Jing*, is probably better known as the *Tao Te Ching*. Pinyin transcriptions such as "Laozi" and "Dao De Jing" will be used in this book. In most places, whenever Chinese characters are provided, traditional Chinese characters will be used, as is more common in books on Classical Chinese. To help the readers better understand the subject matter of this book we have a companion website with useful materials such as recordings, demonstrations of sounds and various exercises. Throughout the book, icons are placed in the margins, where applicable, to refer the reader to relevant materials on the companion website.

The field of Chinese historical linguistics is vast and rich in scholarly debates and disagreements. Therefore, in an attempt to cover all the major topics in the field in such a small book, I have had to simplify many issues and have avoided introducing extensive debates on certain topics. This approach is justified by my belief that, for a book of an introductory nature, it is better for readers to have a good command of the basic methodology and preliminary conclusions than to be exposed to the full range of debates and details. As readers progress to a higher level of research, they will begin to question certain assumptions and simplifications. It will not be too difficult for readers to transition from a simpler picture to further debates and refinements.

Writing a comprehensive history of the Chinese language for the general readership is a daunting task for anyone in the field, but I do feel that there is a need for such a book. Every effort has been made to ensure the accuracy of the content and the citations. Thanks to the invaluable comments from reviewers, the writing has been much improved. The remaining errors and inaccuracies, which I hope will not be many, are certainly mine. If this book proves to be an enjoyable and informative reading experience, that sparks your further interest in linguistic history and the Chinese language and culture, that will be my best reward for having worked day and night in writing this book.

Hongyuan Dong
Washington, DC
September 30, 2013

Acknowledgements

The writing of this book has greatly benefitted from input and help from many people. First and foremost, I would like to thank Professor Young-Key Kim-Renaud of the Department of East Asian Languages and Literatures at the George Washington University for encouraging me to teach two linguistics courses and to write a book. Without her guidance, I would never have envisioned the usefulness of such a book. Students in my Introduction to Chinese Linguistics II class asked many good questions and provided invaluable feedback that helped me to explain the materials in a more accessible way for beginners of the Chinese language and culture and of linguistics. My colleague, Miaochun Wei, a native speaker of Taiwanese, discussed many interesting facts with me regarding the Taiwanese dialect. The comments of the anonymous reviewers have also helped me greatly in improving the accuracy of my writing and I would also like to thank my editors at Routledge – Andrea Hartill took an interest in the topic of this book and encouraged me to make a full proposal and Isabelle Cheng was very accommodating during the process of writing, answering my questions very promptly. I would also like to thank Emma Hudson and Jonathan Merrett for their excellent designing and copyediting expertise that has made this book look its best.

CHAPTER 1

A walk through time

General introduction

"The *Dao* that can be spoken of is not the unchanging *Dao*; the name that can be named is not the unchanging name",[1] goes the opening phrase from the *Dao De Jing*,[2] one of the earliest philosophical works in Ancient China, written by Laozi (*fl.* sixth century BCE).[3] Despite the importance of this book in Chinese culture and thought, its real meaning can be difficult to grasp. Wouldn't it be ideal if one could travel some 2,500 years back in time and have a word with Laozi, although he would probably be very reluctant to explain anything in words? But would a modern speaker of Chinese and Laozi be able to understand each other, even in Chinese? Definitely not! Although the *Dao* is constant in Laozi's philosophy, the language one uses to talk about the *Dao* changes constantly. This book tells the story of how the Chinese language has evolved from the earliest time that we can have knowledge of to its current usage and state in the twenty-first century.

1.1 WHY STUDY LINGUISTIC HISTORY?

One of the major problems of understanding ancient texts is language. If we had a better idea of the language used in Laozi's time, it would be much easier for us to understand his thoughts. The same is true of any historical records. Linguistic knowledge can help us grasp the true meaning of these written records and textual interpretations are also the basis for the study of history and ancient thoughts. Language is an important aspect of human society and a history of language can touch upon how language was used in society at different times. In other words, we can approach history from a linguistic perspective. Even if we only look at how a language has changed, without also looking into the related sociocultural background, it is still a fascinating topic for those of us who love languages. Elements from Classical Chinese can be readily

spotted in Modern Standard Chinese, which makes it necessary for native speakers, researchers into Chinese culture, and students of the Chinese language to have some knowledge of the ancient language.

1.2 THE LINGUISTIC SCOPE OF THE HISTORY OF THE CHINESE LANGUAGE

In what aspects can a language change? There are a certain number of basic sounds in each language. Historical phonology studies how this system of sounds changes. For example, there are 22 consonants in Modern Standard Chinese, but more than 1,000 years ago, in the language that can be said to be the ancestor of Modern Standard Chinese, there were more than 30 consonants. How were they pronounced? How did they change into the 22 consonants of Modern Standard Chinese? Chapters 3 to 5 focus on such changes and try to reconstruct the sound systems at different times. Besides sound change, studies of syntactic change look into how the rules of grammar and forms of grammatical constructions change. For example, "John is scolded by the principal" is a passive sentence. The Chinese equivalent of this construction is formed with the word bèi. Thus we want to know when this usage first appeared in historical texts and how it developed into its current form and usage. Chapters 6–7 will illustrate the key features of Classical Chinese and how some of the typical syntactic constructions in Modern Standard Chinese came about. Another aspect of language change lies in the meaning of words, which will be dealt with briefly in Chapter 8.

We use the term "Modern Standard Chinese" to refer to the official language of China. Chapter 9 will give an historical account of how this national, standard language was created and its major features will also be discussed. Alongside the national language there are many regional varieties of Chinese which, to some extent, developed from the same ancestral language. Chapter 10 will give a brief description of such varieties of Chinese in terms of their connection to the ancestral language. This branch of study is traditionally called dialectology in Chinese linguistics.

This book deals with how the sounds, words and syntax of Chinese

have changed and how the current varieties of Chinese are related to earlier stages of the language. It would probably make our task easier if there were video clips or audio recordings from these ancient times to show how the language was used then, but since we do not have such materials, studies of linguistic history have to rely on written records. The great thing is that written records of Chinese go back at least 3,200 years to the oracle bone script in the Shang Dynasty (sixteenth to eleventh century BCE). But what about the Chinese language before that? Even though we do not have extensive written records from before the Shang Dynasty, the language undoubtedly existed in an earlier spoken form. Using a combination of different methods, we can actually push our hypotheses and knowledge of the language further back to prehistoric times when Chinese was more similar to the ancestral languages of Tibetan and Burmese. We will talk about such genetic relations within the Sino-Tibetan language family in Chapter 2.

Written records are the primary sources of findings in historical linguistics. Actually, the writing system itself should have a history of its own. Chapter 11 will give a sketch of the development of Chinese characters from the oracle bone script to the current simplified forms.

1.3 THE WIDER PICTURE BEYOND THE LINGUISTIC

Although the subject matter of historical linguistics is language itself, it cannot be studied in isolation. When discussing the Chinese language in prehistoric times, we can enlist the help from findings in disciplines such as human population genetics and archaeology. For later stages, where written records are available, the use of these records begins with an inquiry into how and why these written materials were created. Discussions of how language was used in ancient societies can help us better situate our linguistic knowledge, reconstructed through historical linguistics methods. The written records used for the study of the history of the Chinese language include poetry, fiction, Buddhist texts and stories, ancient classics, etc., and these are important cultural products in themselves. To talk about the stages of the history of the language, we will definitely have to touch upon the dynastic history of Ancient China

as well. Generally speaking, language changes mostly by itself, naturally, without conscious human effort. But linguistic reforms can also be carried out by governments and people consciously changing various aspects of a language. Such linguistic reforms are usually motivated by certain sociopolitical factors. Therefore, as we go through the different stages of the Chinese language, we will have to develop some understanding of the relevant cultural, historical, social and political backgrounds.

1.4 TERMINOLOGICAL CLARIFICATION

When the word "language" is used in daily conversation, it can refer to spoken or written forms. But in linguistics, "language" primarily refers to the spoken form. In the context of the title of this book, *A History of the Chinese Language*, what does "the Chinese language" refer to? This is actually a very tricky issue.

First and foremost we have to deal with the more controversial term, "Chinese dialects". Roughly speaking, people in the northern and southwestern parts of China speak Mandarin Chinese. Within this vast area, there are different dialects that are mutually intelligible. For example, a person from the city of Harbin in the Heilongjiang Province in Northeast China can converse with a person from Beijing quite freely if they speak their own dialects. Although it would be a little bit more difficult for a native of Beijing to speak with someone from Chengdu in the Sichuan Province in Southwest China, they can still maintain a high level of mutual intelligibility when they use their own dialects. In the southern and southeastern parts of China, it is often the case that speakers from different regions either have great difficulties understanding each other or cannot communicate at all if they use their local dialects, e.g. Cantonese, Shanghainese, Fukienese, Hakka, etc. A native speaker of Shanghainese cannot communicate with someone who speaks Cantonese, unless they both speak Modern Standard Chinese. Sometimes, even within the same linguistic group, such as Fukienese, it is possible that people from different places cannot communicate with each other in their native tongues. Traditionally, both the mutually intelligible Mandarin dialects and those that are not mutually intelligible

in southern and southeastern China are regarded as fāngyán, which literally means "local speech". The common translation of this term in English is "dialect". But this translation causes a certain degree of misunderstanding and terminological chaos.

According to Mair (1991: 4) the English word "dialect" refers to "one of two or more mutually intelligible varieties of a given language distinguished by vocabulary, idiom and pronunciation." For example, someone from London might speak differently from a Sydney native in terms of pronunciation and vocabulary but they can have a conversation with no major problems. This is true of the different dialects within Mandarin Chinese, but not true of the situation between Mandarin and Cantonese, or between Cantonese and Shanghainese. Mair (1991) compares the situation between these major groups of Chinese that are not mutually intelligible to that between English, Dutch, Swedish, and other Germanic languages. Clearly, the translation of fāngyán as "dialect" is misleading. Thus Mair (1991) proposes a new term "topolect", in which "topo" corresponds to "fāng" and "lect" to "yán". This is indeed a better terminology for what fāngyán means in Chinese. Therefore, it is necessary to keep in mind that the geographic variations of fāngyán have both similarities to, and differences from, the kind of linguistic variations covered under the term "dialect". However, the term "dialect", the traditional translation of fāngyán, has been widely used, or misused, according to Mair (1991) and indeed we have a dilemma with various terminologies. For the lack of a better solution, let's use the term "Chinese dialects" as a convenient way to refer to these different local speeches, while dispensing with the common connotations of the word "dialect". Now that the term "Chinese dialects", or simply "dialects", is established in this book, we can give a more detailed description of the major dialect groups.

Table 1 The major groups of Chinese dialects

Dialects	*English*	*Principal Geographic Distribution*
Guānhuà 官話	Mandarin; Northern dialects	Northern and southwestern China (chiefly north of the Yangtze River)

Wú yǔ 吳語	Wu dialect	Shanghai, Zhejiang and southern Jiangsu
Xiāng yǔ 湘語	Xiang dialect; Hunanese	Hunan
Gàn yǔ 贛語	Gan dialect	Jiangxi
Kèjiā huà 客家話	Kejia dialect; Hakka	Fujian, Guangdong, Guangxi, Taiwan
Mǐn yǔ 閩語	Min dialect	Fujian, Taiwan, Hainan
Yuè yǔ 粵語	Yue dialect; Cantonese	Guangdong, Guangxi, Hong Kong

Traditionally, Chinese dialects are categorized into these seven major groups: Mandarin, Wu, Xiang, Gan, Min, Kejia and Yue. Table 1.1 gives information on the principal geographic distribution of these dialects together with their names in Chinese and in English.

Usually speakers from different groups cannot communicate with each other, although for historical and cultural reasons, the various Mandarin dialects are actually intelligible to varying degrees to the majority of speakers of the other six groups of dialects, but in one direction only. However, if a speaker from the other six major groups was never exposed to Modern Standard Chinese they would still not be able to understand the Mandarin dialects. Mandarin is the largest group in terms of both geographic area and the number of speakers. Each of the other six groups is usually spoken in one or two of the provinces to the south of the region where the Mandarin dialects are spoken. Cantonese is arguably the most familiar dialect for many people outside of China. One reason for this is that Cantonese-speaking people were among the earliest to migrate overseas to North America and Europe.

The fact that there are so many different regional varieties of Chinese calls for a common spoken language that can be used by all speakers,

from all these dialect groups, to communicate with each other. Such a common spoken language can be described as a sort of lingua franca. Historically Mandarin has always been the base of such a lingua franca. In modern times, there is an official, common spoken language called Modern Standard Chinese, or Pǔtōnghuà ("common speech"), which is based on the Mandarin dialect spoken in Beijing. The Mandarin spoken in Taiwan, called Guóyǔ ("national language"), shared the same origin as Pǔtōnghuà although there are noticeable differences in pronunciation, grammar and vocabulary. Thus Guóyǔ should not be confused with Taiwanese, the former being a standard language based on Mandarin and the latter being a Southern Min dialect. Currently Modern Standard Chinese is taught in all schools in China and it is the official language used for publications, radio and TV broadcasts, etc. The promotion of Modern Standard Chinese in the non-Mandarin speaking areas is the main reason why various Mandarin dialects are intelligible to non-Mandarin dialect speakers.

The above-mentioned varieties of the Chinese language are the native tongues of the largest ethnic group in China called the Hàn. There are 55 other ethnic groups in China. Some of these people speak languages that are very different from Chinese, such as Mongolian, Zhuang, Tibetan, Uyghur, etc, while others adopted a variety of Chinese a long time ago. In modern times, people who do not speak a variety of Chinese either learn to speak a Chinese dialect in the process of extensive cultural and economic interactions with local Hàn people, or learn to use Modern Standard Chinese in school.

As we have seen, there are regional varieties of Chinese and an official language which can be used by all speakers to communicate with each other. If we go back in time, we would expect to find a similar situation. In the time of Laozi and Confucius (551–479 BCE), there were dialect differences already, although the differences could not have been as great as they are now. Written records also show that there was a common spoken language at that time. Ever since then we find the same situation in each period of the language. This common spoken language ultimately developed into Modern Standard Chinese. Thus when we say "the Chinese language" in this book, we refer to the common spoken

language in each historical period.

1.5 PERIODIZATION OF THE CHINESE LANGUAGE

Different aspects of the language change at different paces. Usually phonological criteria have been used by many linguists as the basis for the periodization of the Chinese language. If the sound system of the Chinese language remained more or less stable within an extended period of time, then we can regard this timespan as one stage of the language. Table 1.2 shows the major periods of the Chinese language and their relevant timespans, based on Wang (1958) and Xiang (1993).

The prehistoric period preceding Old Chinese can be called Proto-Chinese. Although there are hardly any written records from this prehistoric time, sporadic evidence in Old Chinese helps us deduce some features that might have been present in Proto-Chinese. Comparisons with related languages will also shed light on the properties of Proto-Chinese.

Old Chinese refers to the common spoken language in the first millennium BCE. As has just been mentioned above, in the Spring and Autumn Period and the Warring States Period, there were already different dialects in the various political entities called "states" (guó) in terms of vocabulary and writing systems, and conceivably of pronunciation. Thus, correspondingly, a common spoken form called "the elegant speech" (yǎyán) was used by people from various states to communicate with each other and by educated people to read the classics. This yǎyán is what we call Old Chinese. Language change is gradual – within each period of the language in Table 1.2, cumulative changes make the states of the language noticeably different towards the two ends of the period. Therefore, we can further divide each of the longer periods into sub-stages – Zhengzhang (2003) divides Old Chinese into three sub-stages. The first sub-stage, Early Old Chinese, is the language spoken during the Shang Dynasty and early Zhou Dynasty. The knowledge of the sounds at that time is mostly based on the phonetic cues encoded in Chinese characters found during this early stage, including the oracle bone script and bronze script which was inscribed on bronze ware from the Zhou Dynasty. The middle sub-stage is from the

Zhou Dynasty until the Qin Dynasty. This is actually what most people refer to as Old Chinese proper in terms of the sound system, since the written records on which Old Chinese studies are based are mostly from this period. From the Qin Dynasty until the end of the Han Dynasty is the third sub-stage of Old Chinese, Late Old Chinese. In addition to the sounds of Old Chinese, our knowledge of the grammar and vocabulary of Old Chinese is based on written records produced during the second half of the millennium, mostly the prose writings of the Chinese classics and the major works in the Han Dynasty. The grammar of earlier times, for example in the *Book of Documents* (*Shàngshū*, as early as eleventh century BCE), shows considerable differences from this latter half of the period.

Table 1.2 Periodization of the Chinese language

Periodization	*Timespan and relevant historical periods*	
	Twelfth century BCE to third century CE	
Old Chinese *Shànggǔ Hànyǔ* 上古漢語	Late Shang Western Zhou Spring and Autumn period Warring States period Qin Western and Eastern Han	(12th–11th centuries BCE) (11th century–771 BCE) (770–476 BCE) (475–221 BCE) (221–207 BCE) (206 BCE–220 CE)
	Fourth to twelfth centuries CE	
Middle Chinese *Zhōnggǔ Hànyǔ* 中古漢語	Three Kingdoms Western and Eastern Jin Northern and Southern Dynasties Sui Tang	(220–265) (265–420) (420–589) (581–618) (618–907)

	Five Dynasties	(907–960)
	Northern and Southern Song	(960–1279)
Thirteenth to early twentieth centuries CE		
Early Modern Chinese *Jìndài Hànyǔ* 近代漢語	Yuan	(1271–1368)
	Ming	(1368–1644)
	Qing	(1644–1911)
Early twentieth century (ca. 1911) to now		
Modern Chinese *Xiàndài Hànyǔ* 現代漢語		

Source: Wang Li 1958, Xiang Xi 1993

The Qin Dynasty is regarded as the empire that unified China after a long period of disunity. Qin Shi Huang (259–210 BCE), the first emperor of the Qin Dynasty, carried out systematic efforts to unify his empire. One such measure was to unify the writing system. His troops also reached far into the southern part of the country, bringing Chinese-speaking people into what is the Cantonese-speaking area today.

The subsequent Han Dynasty was a period of major economic and political prosperity when the ethnic identity of the Chinese people was solidified. In contrast to the non-Chinese peoples outside the Han Dynasty, the Chinese people considered themselves the Hàn people. Hence one of the names of the Chinese language is *hànyǔ*, meaning the language of the Hàn people. It is also during the Western Han Dynasty that Confucianism was established as the official ideology of the empire. The significance of this for the study of the language lies in the fact that Confucian classics were studied almost religiously by every scholar up to the early twentieth century. Many of these Confucian scholars noted the

differences between their own spoken language and the language recorded in the Confucian classics. Such sporadic records can be valuable materials. During the Eastern Han Dynasty, Buddhism was introduced into China. Translations of Buddhist canons during this time also provide valuable data as to how the Chinese language was spoken, which can be revealed from a comparison of the original Sanskrit terms and the terms in Chinese, especially if the term was translated based on the correspondence of sounds between the two languages.

Between the Eastern Han Dynasty and the Sui Dynasty (581–618 CE), there was a long period of political division. Constant wars in the northern part of the country drove many Chinese-speaking people to the south, while non-Chinese speaking people ruled the north. Also during this period of domestic turmoil, the sound systems of Old Chinese gradually transitioned into a more complex system by the beginning of the Sui Dynasty as recorded in a rime dictionary called *Qièyùn* written during the late sixth and early seventh centuries. The sounds of the *Qièyùn* dictionary are mostly regarded as the literary pronunciation of the Northern and Southern Dynasties. Within the grammar and vocabulary system we can trace the origins of many elements in various modern Chinese dialects up to this period as well. By the end of this transitional period what can be considered Middle Chinese was already formed. Middle Chinese is important in that it can be regarded as the common ancestral language for most modern Chinese dialects. Therefore, knowledge of Middle Chinese is the key in understanding how the modern spoken varieties of Chinese came about.

The Imperial Examination system was established in the Sui Dynasty in 605 CE as a way of selecting qualified Confucian scholars to serve as administrative officials in the imperial government. The system was used in subsequent dynasties up until the early twentieth century. This examination system gave rise to official dictionaries, which are similar to the *Qièyùn*, compiled to set the correct pronunciation for Chinese characters used in the composition of essays and poetry which were usually part of the examination. Such rime dictionaries are valuable sources of the sound systems of different times.

The Tang Dynasty is known in Chinese history as a major empire with

great cultural and historical significance. In terms of the literature, Tang poetry has always been upheld as the highest achievement of the literary genre. Great poets such as Li Bai (701–762 CE) and Du Fu (712–770 CE) are household names in all Chinese-speaking areas. Almost all Chinese speakers can recite a number of Tang poems from memory. Actually Tang poetry contains important clues to how the literary language was spoken during that time and studies of Middle Chinese are thus intertwined with Tang poetry and poetics.

After the Tang Dynasty, there was a relatively short period of political disunity, but soon the country was unified by the Northern Song Dynasty. By this time, the main phonological features of the earlier part of Middle Chinese, as represented by the *Qièyùn*, had changed quite a bit and Middle Chinese was in the process of transitioning into Early Modern Chinese. A special type of written material from the Song Dynasty, called the rime tables (yùntú), based on the sound system of the Qièyùn dictionary, is an important source which can help us understand the sounds towards the end of the Tang Dynasty and in the Song Dynasty; it is the link between the earlier part of Middle Chinese and the next major stage of the Chinese language. Pulleyblank (1970, 1971) regarded the language represented by the rime tables as Late Middle Chinese, in comparison to the Early Middle Chinese of the *Qièyùn* time. Due to the transitioning status of the Chinese language during the Song Dynasty, there are still debates about where the Song Dynasty should belong in the periodization of the language. Jiang (2005: 5) points out that, at least in terms of grammar and vocabulary, what we can call Early Modern Chinese had been established by the beginning of the Song Dynasty. In terms of the sounds, Norman (1988) thinks that the language of the Song Dynasty should be considered an earlier form of Old Mandarin, i.e. Early Modern Chinese in Table 1.2.

A rime book called *The Phonology of the Central Plains* (*Zhōngyuán Yīnyùn*), written in 1324 in the Yuan Dynasty, is the major source of what is referred to as the sound system of Early Modern Chinese. Major changes had taken place in both the consonants and vowels of Middle Chinese. Many properties of Modern Standard Chinese can be already found in the system of this rime book. This can be regarded as the first

stage of Early Modern Chinese. The second stage includes the Ming and Qing Dynasties, when important moves towards becoming Modern Standard Chinese, especially in the consonants, took place. Generally speaking, by the middle of the Qing Dynasty, the common spoken language was already sufficiently similar to Modern Standard Chinese in all aspects including the sounds, vocabulary and grammar. There are novels written during this time in a style that is close to the actual spoken language.[4] For speakers of Modern Standard Chinese, these literary works are quite accessible and easy to read.

Starting from the late Qing Dynasty, towards the end of the nineteenth century, progressive scholars looked to the West for ways of modernizing China, among which building a national language became a priority. In 1911, the Qing Dynasty was overturned, ending more than 2,000 years of rule of imperial dynasties in Chinese history, bringing China into modern times. In the Republic of China, the standard for a national language continued to be debated. The Beijing dialect was chosen in 1913 as the basis of the new national language, although at that time there was still disagreement on whether elements from other dialects should be included. After long, heated debates over the next decade, it was finally agreed upon among the scholars involved that the new standard of pronunciation should be solely that of Beijing. In 1949, the People's Republic of China was founded on the mainland. The government implemented reforms on multiple levels to promote a national language called Pǔtōnghuà (literally "common speech") and a Romanization system called "pinyin", used primarily as a way to annotate the pronunciation of Chinese characters. The promotion of Pǔtōnghuà in China has been quite successful. Now, after a few decades of promotion, the majority of people from different dialect regions in China can understand and speak Pǔtōnghuà. At the same time, various Chinese dialects are affected by Pǔtōnghuà to different degrees, because younger generations are taught to speak Pǔtōnghuà in school and modern technology and media have a homogenizing effect, bringing Pǔtōnghuà to every corner of the country. In recent years, there has been a rise in the awareness of regional cultures and identities that are deeply rooted in the use of different local speech. Thus a balance is yet to be established between promoting a uniform national language and preserving the

diversity of regional varieties of Chinese, regional cultures and identities.

This chapter gives a very brief sketch of the history of the Chinese language from prehistoric times to its current usages. The remainder of this book will be devoted to a more detailed look at each stage of the development of the Chinese language.

NOTES

1 This quotation is based on the translations of James Legge (see Lao-tse 1891 in the bibliography) and D. C. Lau (see Lau Tzu 1964).

2 All Chinese terms are transcribed in pinyin. Please refer to Appendix III and Appendix IV for more information on the pinyin notation.

3 There is still debate among scholars as to whether the work *Dao De Jing* was really written by one person, and whether there was such a person called Laozi.

4 Literary Chinese, or to put it in a less accurate way, Classical Chinese, was still used at this time for formal writing and the so-called "high-brow" literary genres such as poetry and essays.

CHAPTER 2

Where it all began

Prehistory

One day, in the early 1780s in Calcutta, India, as Sir William Jones was studying the ancient language Sanskrit, he became more and more intrigued by how Sanskrit verbs and grammar had a strong affinity to those in Greek and Latin. "This couldn't possibly have been a coincidence," he thought. The discovery led to his view that these three ancient languages had a common source. He hypothesized a common origin, "which, perhaps, no longer exists", for languages such as Sanskrit, Greek, Latin, Old Persian, Gothic and Celtic. This common origin is what was to be called the Indo-European language, which has now been successfully reconstructed in great detail.

2.1 ESTABLISHING LINGUISTIC GENETIC RELATIONSHIPS

One of the most obvious ways that languages sound similar lies in how words sound alike among them. Table 2.1 shows a few words from Sanskrit, Greek and Latin to show how these languages are similar.

But, of course, words can be similar for different reasons. First, they can be similar by accident. The Classical Chinese word "dàn" means more or less the same thing as "dawn" in English. Such similarities are not systematic, since we can only find a few such words. Second, words can be similar because they mimic natural sounds. Cats "miāo" in Chinese and they "meow" in English. Onomatopoeic words can be similar across many different languages. Third, words can be borrowed from one language into another with corresponding phonological adaptations. It is probably not difficult to guess what the Chinese word "màikèfēng" means, since it is the word "microphone" borrowed from English.

Generally speaking, what sounds are used to represent what words in a language is, for the most part, arbitrary (Saussure 1983). While people say "water" in English, the Chinese use the word "shuǐ". They are totally different because the association between sounds and words is arbitrary. However, when we see that the Germans say "Wasser", we have reason to suspect that "Wasser" and "water" might have come from the same word, also because of the arbitrariness of the association between sounds and words. Such genetic similarities should be quite extensive and systematic in that there are many such words and the sounds in one language correspond to those in another in a regular fashion. Similarities due to coincidence or sound imitation are not extensive, or regular. Loanwords can be fairly extensive if a large number of words are borrowed between two languages, but if we have evidence for the borrowing, then we can exclude the possibility of similarity between these languages due to a common origin. Therefore, the first step in establishing a genetic relationship between two languages is to compile a list of cognate words, such as those in Table 2.1.

Once we have such a list, it is necessary to look at sound correspondences, rather than word shape similarities. For example, in Table 2.1, in the first word, we find that the sound p- in Sanskrit corresponds to p- in Greek, p- in Latin, and f- in English. Thus a p-p-p-f correspondence between these languages is obtained. Actually we can find the same sound correspondence in the third word meaning "foot" as well. If a longer list of cognate words is available, this same correspondence can be found in many cognate words. Now let's look at the second word: the d- in Sanskrit corresponds to d- in Greek, d- in Latin, and t- in English. Thus a d-d-d-t correspondence is established. The same correspondence can be found in the third word as well. Again if we have a longer list, we can indeed find many more examples of such correspondences. Therefore there are regular sound correspondences between these languages. Thus we have finished the second step towards establishing a genetic relationship, i.e. finding regular correspondences.

The third step will be the reconstruction of the ancestral language called the proto language. In the case of Indo-European, the reconstructed ancestral language is called Proto-Indo-European. The

main purpose of the reconstructed system is to explain how the ancestral language developed into different languages. We will pick up the third step in a litile bit of more detail in Chapter 4, when we talk about the reconstruction of the sound system of Middle Chinese.

So far we have seen evidence for the genetic relatedness of Sanskrit, Greek and Latin. Actually English is genetically related to these languages, but since the connection between English and the other three languages is more indirect, let's focus on the ancient languages for now. Figure 2.1 shows the genetic relationship between Sanskrit, Greek, and Latin in a structure that looks like a family tree.

The common ancestral language is called Indo-European. Through some changes, or innovations, in the sound system of Proto-Indo-European, each of the three languages became a different language. Each of these new languages can be further split into more daughter languages; for example, Hindi, Bengali, and Marathi developed from Sanskrit. This representation, and how we understand language change, is called the Family Tree Model. Figure 2.1 is a schematic representation of the Indo-European family of languages; in fact there are over 400 languages and dialects in the whole family, only a small portion of which is sketched here.

Table 2.1 Indo-European cognate words

Sanskrit	*Greek*	*Latin*	*English*
pitar-	pater	pater	father
dva-	dyo-	duo-	two
pad-	pod-	ped-	foot
aksah	haksos	axis	axis

Source: based on Xu 1991 and Campbell 1999.

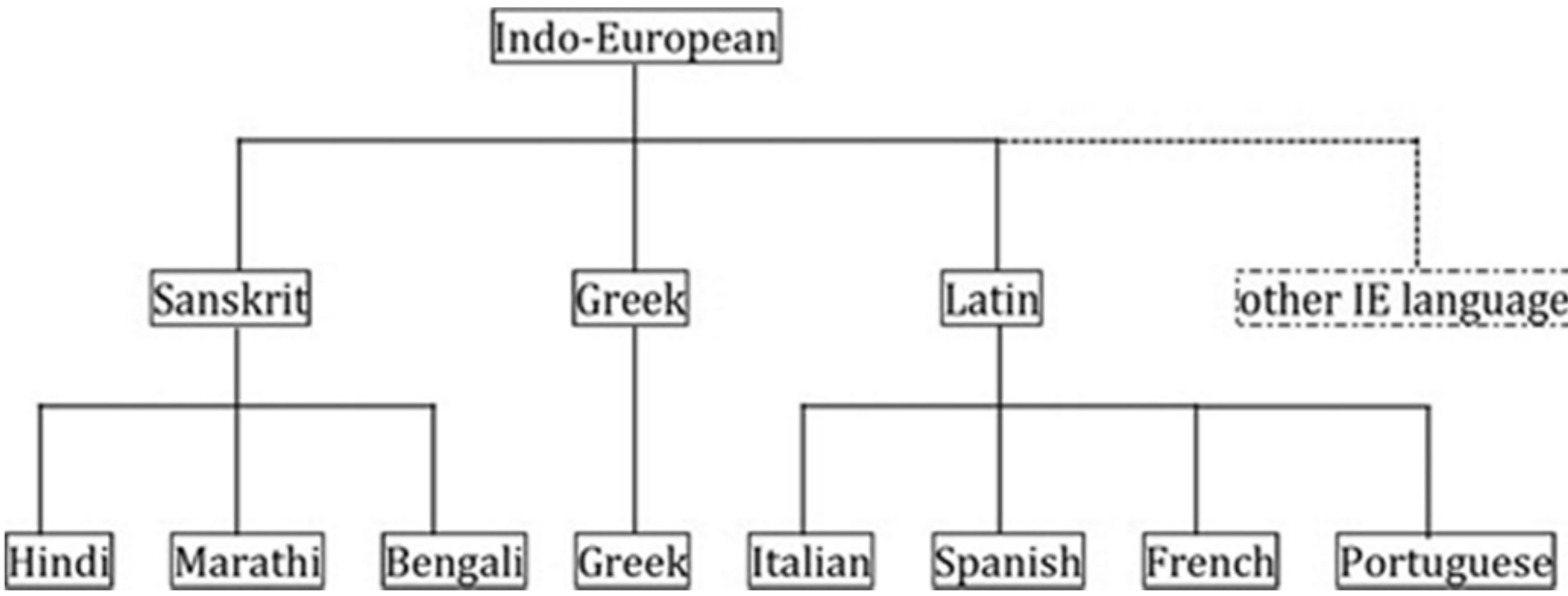

Figure 2.1 A schematic family tree of Sanskrit, Greek and Latin

2.2 THE SINO-TIBETAN LANGUAGES

After the concept of the "language family" had been established in the early nineteenth century, based upon the Indo-European model, linguists started to look at other languages and tried to classify them into different families. What other languages are genetically related to Chinese? We may look at those languages that are geographically neighbors of the Chinese language.

To the north, northeast and northwest of Chinese, we find languages such as Mongolian, Manchu, Japanese, Korean and Uyghur. These languages belong to the Altaic family, although some people believe that Japanese and Korean are isolates, meaning that they are not related to any other language or to each other. A quick look at Japanese and Korean words may reveal a high degree of similarity between these two languages and Chinese in terms of the vocabulary. To the south of China, we find languages like Vietnamese, which belongs to the Austroasiatic family. There are also a large number of words in Vietnamese that are similar to those in Chinese. Table 2.2 shows examples of numerals from Mandarin (Modern Standard Chinese), Cantonese, Japanese, Korean and Vietnamese.

There are two sets of pronunciations in the columns for Japanese, Korean, and Vietnamese. The first pronunciation for each word looks similar to Mandarin and Cantonese. However, the second pronunciation does not look like Chinese at all. Therefore, it is very likely that the

second pronunciation is the native pronunciation in these languages, while the first one was an early loanword pronunciation. Historically speaking, Classical Chinese and Chinese characters were borrowed into Japan, Korea and Vietnam during the Middle Chinese period. The pronunciation of Chinese characters was also imported into these languages with corresponding phonological adaptions. Therefore there is no genetic relationship between Chinese and the neighboring languages of Japanese, Korean and Vietnamese.

Table 2.2 Numerals in Chinese, Japanese, Korean and Vietnamese

English	*Mandarin (Pinyin)*	*Cantonese (Jyutping)*	*Japanese*	*Korean*	*Vietnamese*
one	yī	jat[1]	ichi/hito	il/hana	nhất/một
two	èr	ji[6]	ni/futa	i/dul	nhị /hai
three	sān	saam[1]	san/mi	sam/set	tam/ba
four	sì	sei[3]	shi/yon	sa/net	tứ/bốn
five	wǔ	ng[5]	go/itsu	o/daseot	ngũ/năm

In more modern times, around the early part of the twentieth century, especially during the New Culture Movement when Chinese intellectuals looked to modernize China, many words related to Western culture, science and technology were borrowed into Chinese from Japanese. A large number of words from major European languages were translated into Japanese using kanji characters in the second half of the nineteenth century as a result of the Meiji Restoration that started in 1868. For example, words like “kēxué” (“science”), “mínzhǔ” (“democracy”), “shèhuì” (“society”), “jīngjì” (“economy”), “diànhuà” (“telephone”), etc. have all been borrowed from Japanese in modern times.

For the remaining major Altaic languages mentioned above, there are also a smaller number of loanwords between Chinese, Mongolian and Manchu. For example, the word “zhàn” (“a station, a stop”) currently used for a bus stop, subway stop or a train station was originally borrowed from Mongolian. A number of colloquialisms in Beijing Mandarin and northeastern Mandarin were borrowed from Manchu

during the Qing Dynasty as well, for example, “mǎmǎhūhū” (“in a careless way”). There are also a number of Chinese loanwords in Mongolian and Manchu, due to close cultural contacts over a period of time. Apart from loanwords, there is no list of cognate words that can be found to establish a genetic relationship between the Altaic languages and Chinese, or between the Austroasiatic languages, e.g. Vietnamese and Chinese.

Also relevant here is a language family called Austronesian, which includes the Formosan languages spoken in Taiwan and also the languages spoken across the vast area between Taiwan and Indonesia, such as the Filipino and Indonesian languages. If there are connections between Chinese and Austronesian, such connections have to be quite remote. It is possible for the Austronesian people to have originated in southern China before crossing the Taiwan Strait into the island and then migrating further to other areas between Taiwan and Indonesia, as pointed out by Wang (1998). Sagart (1993) proposed a genetic relationship between Chinese and Austronesian, but generally linguists have not been able to agree upon a conclusion in this area.

In some parts of southern and southwestern China and in Southeast Asia we find Zhuang, Dong, Thai and other related languages that belong to the Tai-Kadai family. The subgroup of the Tai-Kadai family that includes the Zhuang, Dong and Thai languages is called Kam-Tai, or Zhuang-Dong in China. In southwestern and western China and the Himalayan regions we find languages like Tibetan, Burmese, Qiang and rGyalrong, which are called Tibeto-Burman. There are also the Miao and Yao languages that are spoken in Hunan, Hubei, Guangxi and southwestern China, which belong to the Hmong-Mien family, or Miao-Yao in China. There are a large number of similar words between Chinese and the Kam-Tai, Hmong-Mien and Tibeto-Burman languages. Unlike the similar words between Chinese, Japanese, Korean and Vietnamese, which are known to be loanwords, we do not have readily available records to prove whether the similar words between Chinese and the Kam-Tai, Tibeto-Burman and Hmong-Mien languages are loanwords or true cognate words. Therefore, some linguists call these words “relational words” (“guānxìcí”) until strong evidence can be

produced to prove either way. If the relational words are loanwords, then there would be no genetic relation between the relevant languages. If the relational words are indeed true cognates, then a genetic relationship could be established.

The concept of a Sino-Tibetan language family became popular during the 1930s, at the University of California, Berkeley, because of the work published by the Sino-Tibetan Philology Project. According to Li (1973), based on a paper written by him in 1937, the Sino-Tibetan family includes Chinese, Kam-Tai, Hmong-Mien, and Tibeto-Burman. Benedict (1972), in a paper which was actually written in 1942, argued that the Hmong-Mien languages and the Kam-Tai languages are not part of the Sino-Tibetan family because the relational words between Hmong-Mien, Kam-Tai and Chinese could possibly be very early borrowings.

A basic vocabulary is needed in order to compare words in different languages. Swadesh (1971) proposed a list of 100 words which were meant to be "universal" and stable; since they refer to fundamental and essential aspects of life and culture, these words can be found in every language and they are less likely to be borrowed or replaced. For example, words like "I, you, one, two, man, woman, fish, dog, ear, eye, see, hear" are among these basic words. The 100-word list was compiled from an earlier 200-word list also proposed by Swadesh. The 100-word list comprises those words that are the more basic ones among the 200 words. Generally speaking, higher-level cultural vocabulary items are easily borrowed through cultural contacts, while words on the 100-word list are less prone to be borrowed and, if they do change over time, the change takes a much longer time than with higher-level cultural words.

Since many of the relational words in the Sino-Tibetan family are from such basic vocabulary items, it has long been quite controversial as to how to prove that either these words were loanwords or they were true cognates. Although we do not know the details of linguistic contacts in the past, we can study the process of such contacts in modern times and try to see how borrowing between two languages works in order to understand whether or not relational words are indeed the results of borrowing that happened in the past. Chen (1996) studied how the people who speak the Dai language, a Kam-Tai language spoken in

Southwestern China, and the Chinese-speaking people interact with each other and learned each other's languages. He spent several years documenting the details of linguistic contacts between these two languages, including loanwords. Based on his observations, Chen (1996) proposed a new theory arguing that loanwords can emerge from basic vocabulary, according to a hierarchy, and that there is no boundary as to how deep such borrowings can go. Chen (1993) divides the 100 basic words into three groups. Group A is the most basic, including words like "fly, dog, yellow, hand, sun, I, rain". Group B includes "white, nose, egg, ear, you, bird, new, know". Group C includes "not, two, drink, gray, name, man, woman, what, small, swim". These groups form a hierarchy among the basic 100 words. Using a vast amount of data about known true cognates and known borrowings from different languages, he discovered that when two languages share many relational words due to contact, these shared words tend to be the less basic ones, those that are higher up the hierarchical structure of basic vocabulary, and shared words in the more basic vocabulary tend to be fewer. On the other hand, if two languages share a lot of relational words due to a common origin, e.g. among the different Chinese dialects, or among the Germanic languages including English, German, Dutch and Swedish, these shared words tend to be the more basic vocabulary items, while the less basic ones tend to be fewer. Therefore Chen (1993) suggests that the tendency holds across different languages, and that it has the property of being a generalizable theory.

Based on this theory, Chen (1993) calculated the ratios of the relational words among Chinese, Kam-Tai, Hmong-Mien and Tibeto-Burman. It turns out that Chinese and Kam-Tai share fewer words on the lower hierarchical level, but more on the higher level. Thus, if his theory is correct, the computational results suggest that Chinese and Kam-Tai are not genetically related. The relational words between them are the result of deep cultural contacts and sharing. Chen (1996) expanded the word list to the 200 basic words. In similar fashion to the previous studies mentioned above, he divided the 200 words into two groups: Group 1 included the first 100 words which are regarded as more basic; Group 2 included the second 100 words. He then applied the same mathematical approach to languages that share relational words, and the

conclusions are the same as before. The computational results from Chen (1996) show that Chinese and Tibeto-Burman share more words on the lower hierarchical level, but fewer on the higher level. This clearly suggests that Chinese and the Tibeto-Burman languages are genetically related. At the same time, the Hmong-Mien languages have also been generally excluded from the Sino-Tibetan family by most linguists.

Therefore, of the two main proposed descriptions of the membership of the Sino-Tibetan family mentioned above, Li (1973) and Benedict (1972), more linguists are inclined to agree with Benedict (1972), and include just Chinese and the Tibeto-Burman languages. Table 2.3 shows a few true cognate words in Chinese, Written Tibetan and Written Burmese. The reconstructed pronunciations from Old Chinese are also given since the sound correspondences between Old Chinese and the Tibeto-Burman languages are more easily spotted. The words are also given in Chinese characters and their pronunciation in Modern Standard Chinese.

Table 2.3 Sino-Tibetan cognate words

Words	*Modern Standard Chinese*	*Old Chinese*	*Written Tibetan*	*Written Burmese*
eye	mù 目	mug	mig	myak
six	liù 六	rug	drug	khrok
insect	fú 蜉	bu	’bu	pûi
poison	dú 毒	duug	dug	tok

Source: based on Norman 1988; Zhengzhang 2003

We have pointed out earlier in this chapter that mere similarity in word shapes is not enough to establish a genetic relationship between two languages. What is more convincing is regular sound correspondence. In Table 2.3, the final consonant in the first word meaning “eye” is -g in Old Chinese, -g in Written Tibetan, and -k in Written Burmese. Thus we have a -g-g-k sound correspondence. This same correspondence can be easily seen in the second word “six” and the fourth word “poison”. Also the vowel in the second word “six” is -u in Old Chinese, -u in Written

Tibetan, and -o in Written Burmese. Thus we have a u-u-o correspondence. It is also shown to some extent in the fourth word "poison" as well, although the vowel in Old Chinese is a long vowel, -uu, in "poison" according to the reconstructions of Zhengzhang (2003). Figure 2.2 shows a schematic representation of the Sino-Tibetan family tree.

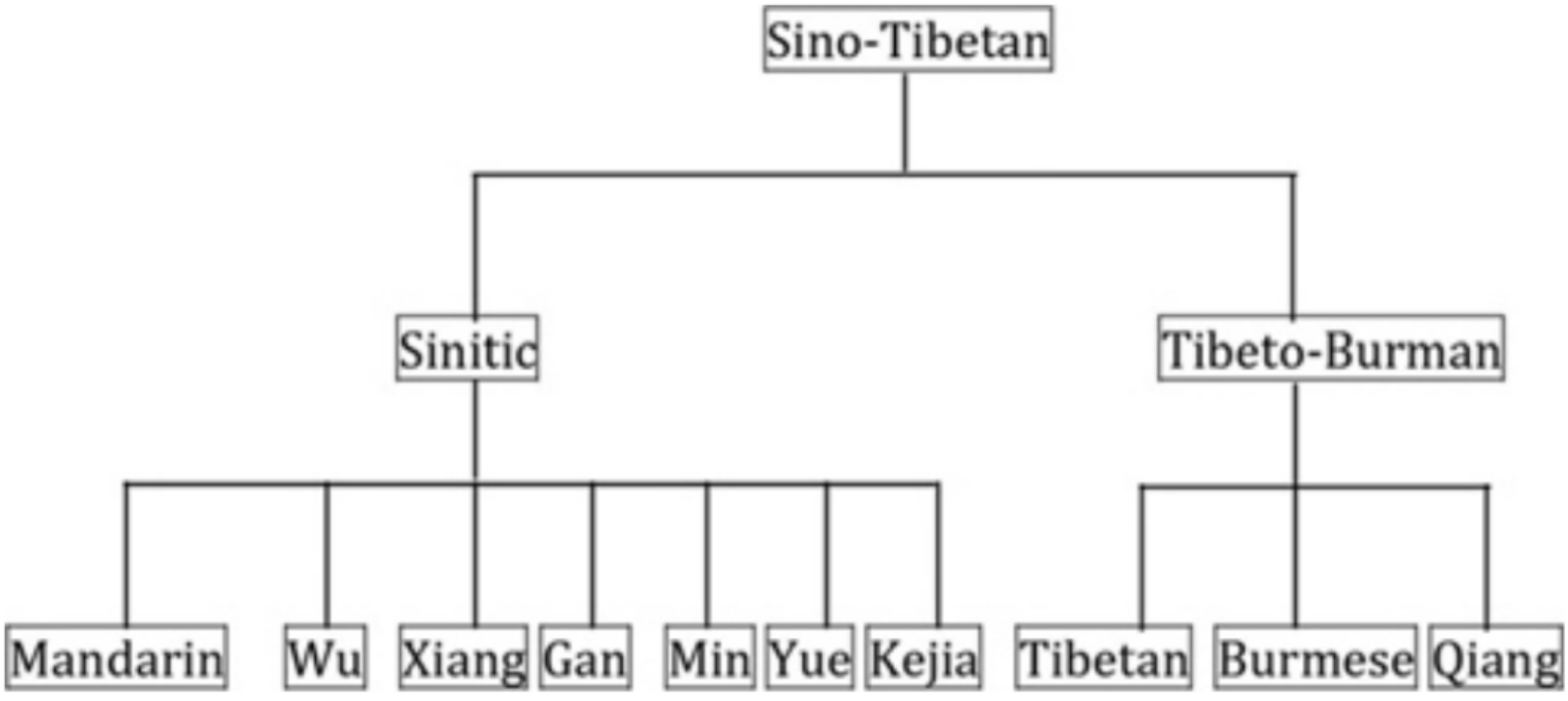

Figure 2.2 A schematic family tree of Sino-Tibetan

In this family tree the primary distinction is between Chinese, or the Sinitic branch, and Tibeto-Burman. The Chinese branch can be further split into the major Chinese dialects mentioned in Chapter 1. The Tibeto-Burman branch includes Tibetan and Burmese amongst other languages.

2.3 HOW OLD IS THE SINO-TIBETAN LANGUAGE FAMILY?

Now that we have established the concept of a Sino-Tibetan language family, how old is this family and when did such a Sino-Tibetan language exist? Earlier we introduced the list of 100 basic words. These basic words are said to change or to be replaced in any language at a rather slow rate when compared with higher-level cultural words. Swadesh was inspired by the method of carbon-14 dating, used to determine the age of fossils of animals or plants. Carbon-14 is a radioactive isotope of carbon which has a half-life of about 5,700 years. If we assume that the 100 basic words in languages also change at a

constant rate, and also that no contacts have been made between the two daughter languages after they split, then by studying the percentage of shared cognate words in two languages we can calculate the time when the ancestral language split in two. This is definitely a marvelous idea, but it has proved to be inaccurate. More recently, mathematical methods developed in biology have been applied to the study of linguistic prehistory. The basic idea is based on the distance between languages, a notion that is related to the percentage of shared cognates. Wang (1994), using data on the distance between major Chinese dialects derived from Xu's (1991) shared cognates data, demonstrated how a tree can be drawn to show the relative distance between any pair of languages by the length of the shortest path connecting these two languages. In Figure 2.2, showing the family tree of the Sino-Tibetan languages, the lengths of the branches are equal, but in Wang's (1994) tree, the lengths are different according to the distance numbers. Wang (1998) further shows how the method can be used to study the age of the Sino-Tibetan family.

First, Wang (1998) collected data on distances between the following three groups of languages: major Chinese dialects, major Sino-Tibetan languages, and representative Indo-European languages including Singhalese, Tajik, Portuguese and such Germanic languages as English, Danish and German. Then three trees were drawn in accordance with the distance matrices for each group. They were then compared and it was shown that the time-depth of the Chinese dialects is comparable to that of the Germanic languages. This corresponds well to the fact that the Chinese dialects began to develop about 2,000 years ago, roughly at the same time that the Germanic languages started to split. This is good evidence of the usefulness and accuracy of this method. The Germanic languages form a sub-tree of the whole Indo-European tree. Since the time-depth of the Germanic sub-tree is about 2,000 years, we can now compare the length of the Germanic sub-tree to that of the Indo-European. It shows that the length of the Indo-European tree is about 3.5 times that of the Germanic sub-tree. Now we can tentatively say that the Indo-European language is about 7,000 years old. Wang (1998) points out that there are actually two major theories for when the Indo-European language started to split: 6000 BP and 8500 BP. Therefore, the 7000 BP here situates right between these two numbers. A comparison

between the tree for the Chinese dialects and the Sino-Tibetan tree shows that the latter is about three times as long as the former. Therefore, it seems to suggest that the Sino-Tibetan language is 6,000 years old, a little bit younger than the Indo-European language. If this is correct, we should be able to find archaeological and even genetic evidence to support such a claim.

Wang (1998) cites three maps, constructed by archaeologists, from Chang (1986) showing prehistoric cultures in East Asia at 9000 BP, 7000 BP and 6000/5000 BP. The map at 9000 BP shows several cultures in the upper and middle Yellow River region, and some in the lower Yangtze River region and southern China. These cultures showed no clear evidence of interactions. The map for 7000 BP contains three major cultures that span the whole region of northern China, two cultures in the lower Yangtze River and a vast culture that spans the southernmost coastal area of China. At this stage there were still no major interactions between these cultures. The third map for 6000/5000 BP presents a picture of a network of cultures from north to south that had a significant degree of interaction. Wang (1998) concludes that population movements brought about similarities between these cultures and, alongside such migrations, people brought their own languages. Therefore, a former linguistic community can split up into different linguistic communities, while formerly different linguistic communities can come into close contact with each other and may borrow words from each other. A possible scenario that can be reconstructed from the above-mentioned maps of prehistoric cultures is that the Sino-Tibetan people originally inhabited the Upper and Middle Yellow River region, and then some of them migrated down the river to the east and then to the south, coming into contact with other cultures. This scenario is supported by linguistic evidence.

In the first millennium BCE, i.e. the Old Chinese period, there were various non-Sinitic peoples in the lower Yangtze River region and southern China. According to historical records, these peoples were called the Bǎiyuè, i.e. Hundred Yue. They had very different cultures from the northern Huáxià culture, i.e. what was to become the Chinese culture. As for the language that the Baiyue peoples spoke, they could be

related to the Austroasiatic, Kam-Tai and Hmong-mien languages that were mentioned earlier. If the ancestors of these languages were indeed those that were spoken by the Baiyue, we may find very early loanwords in both languages as a result of cultural contacts.

Let's look at some examples of such loanwords now. There are two major words that can be used to refer to rivers in Chinese: jiāng and hé. They are used more or less interchangeably to mean "river" in Modern Standard Chinese, although jiāng is usually a major river. Normally jiāng cannot be described by prefixing the word xiǎo ("small"), but hé can be prefixed with either dà ("big") or xiǎo ("small"). Originally jiāng is a proper name referring to the Yangtze River, and hé is a proper name for the Yellow River. In ancient times, the Chinese people inhabited the Yellow River region in the north while a group of the Yue people lived along the lower Yangtze River. Therefore these two words were how these different peoples referred to the rivers of their homelands. Later these two words would eventually be generalized to refer to rivers. The word hé might be a native Chinese word that is related to the Written Tibetan word "rgal-ba" for "to pass or ford a river" (Schuessler 2007; Coblin 1986). Norman (1988) relates the word jiāng to Austroasiatic languages such as Vietnamese. The reconstructed pronunciation of jiāng in Old Chinese is *krung; the reconstructed pronunciation in Vietnamese is *krong. The name of the Yangtze might have been given by the earlier inhabitants when the Chinese people arrived there, adopting the name.

Norman (1988) also gives an example of a loanword from the ancestral language of the Hmong-Mien language. According to him, there are two words for "dog" in Chinese: quǎn and gǒu. The first one, quǎn, cannot be used as an independent word in most Chinese dialects now, except in some Min dialects. It is preserved mostly as part of a larger word. There are clear Sino-Tibetan cognates of this word, showing that it might be the original Chinese word for "dog". In most Modern Chinese dialects, the word for "dog" is gǒu. Norman (1998) cites Purnell's (1970) reconstruction of proto-Miao-Yao, i.e. Hmong-Mien, in which the word for "dog" is *klu^{2}. The "u^{2}" part means that it is some type of "u" sound although the exact phonetic value is yet to be decided. According to Zhengzhang's (2003) reconstruction of Old Chinese, the

Old Chinese pronunciation of “gǒu” is *koo?. The similarity between these two words can also be established with some imagination. Therefore it can be evidence of early contacts between the Chinese people and the Hmong-Mien people in the south.

Chen (1993) presents data for the contacts between the Chinese people and the Kam-Tai people. There has been archaeological evidence for the early cultivation of rice in southern China. The word for rice plant in Classical Chinese is hé, written with a different character from the word for “river” mentioned above. In Zhengzhang’s (2003) reconstruction, the pronunciation in Old Chinese of hé is *gool. Chen (1993) listed the following pronunciations: hau^{4} (Wuming Zhuang) and kha^{2} (Thai). Thus it is possible that the word for rice plant was borrowed into Chinese when the Chinese people got into contact with the Yue people in the south and learned how to cultivate rice from them.

Therefore, the evidence from linguistics and archaeology supports a migration scenario whereby the Sino-Tibetan people moved from their homeland in the present-day Shaanxi and Ningxia area to the east and then south, getting into contact with the Yue people who spoke languages which can be considered the ancestors of modern Austroasiatic, Hmong-Mien and Kam-Tai languages. There was also a separate migration from the Sino-Tibetan homeland further west and then south to Tibet and the Himalayan region. We have not produced concrete archaeological evidence for this migration, although linguists have established over 300 cognate words between Chinese and Tibeto-Burman languages, but we might look to human population genetics for evidence of this side of the story.

As Wang (1998) points out, genetics and languages often go separate ways. For example, during the Qing Dynasty, many Manchu people adopted Mandarin Chinese as their native language. Currently, most of the Manchu people in China can only speak Chinese, while the Manchu language is spoken by the Xibe people in Xinjiang and also by Manchu people in a few villages in Northeastern China. The Manchu language is an Altaic language which does not have any genetic relationship with the Chinese language. However, recent research in human population genetics can use a more reliable genetic marker to trace human

migration. Su *et al.* (2000) says "as delineating migrations becomes one of the major themes in human evolution studies, Y chromosome markers began to show their power in tracing human prehistory." Their studies analyzed the genetic structure of 31 Sino-Tibetan populations represented by 607 individuals residing in East, Southeast and South Asia. They concluded that the Sino-Tibetan homeland is the upper and middle Yellow River basin and about 5,000–6,000 years ago the Sino-Tibetan language began to split into the Sinitic branch and the Tibeto-Burman branch. The Tibeto-Burman group migrated westwards and then southward to the Himalayas. The Sinitic group expanded to the east and south.

2.4 LINGUISTIC PROPERTIES OF PROTO-SINO-TIBETAN

Now one more question is in order. What was this Sino-Tibetan language like? The reconstruction of Proto-Sino-Tibetan has not yet been carried out to the same extent as Proto-Indo-European. We cannot give a full picture of the sounds, words, and grammar of Proto-Sino-Tibetan yet and can only look at some typological features of Chinese and Tibetan in the hope of guessing what the characteristics of the original ancestral language were.

The branch of linguistics that classifies languages into different types based on their formal features, rather than genetic relationship, is called typology. One of the most basic criteria for classification is the basic word order type. For example, in the English sentence "John loves Jane", "John" is the subject, "loves" is the verb and "Jane" is the object. This sentence has the Subject-Verb-Object, or SVO, word order and, typically, sentences in English are in the SVO form. Thus we say that English is an SVO language. Modern Chinese is normally considered an SVO language, while Tibetan is SOV. Another feature of Chinese is that there are very few morphological changes to word shapes, while Tibetan has a complex morphological system. In terms of sounds, Modern Standard Chinese has four tones. Modern Standard Tibetan, based on Lhasa, has four tones, while Amdo Tibetan, spoken in the Qinghai Province region, does not have tones. Tibetan allows for more than one

consonant at the beginning of a syllable, e.g. zgo ("door") and pki ("carry") as in Daofu Tibetan (Hu 1980). Consonant clusters such as zg- and pk-in these words are not allowed in Modern Standard Chinese. Only one consonant is allowed at the beginning of a syllable.

Note that typology classifies languages according to their similarities in form, without regard to their genetic relationship. In fact, typologically similar languages often belong to many different language families. As we mentioned above, English and Chinese can both be argued to be SVO languages, but they are not genetically related. Here, however, with respect to the Sino-Tibetan family, it has been established that languages like Tibetan and Burmese are genetically related to Chinese. Therefore, if these languages are typologically very different now, it is possible that the proto-language had some of the typological features of the modern languages, although it is equally possible that many of these modern typological features were later developments after the proto-language split into different languages. Nonetheless, looking at the linguistic features of Tibetan, Burmese and Chinese we can still, to some extent, try to figure out what features the proto-language had.

Looking at all the different parameters that could be incorporated into the yet-to-be reconstructed Proto-Sino-Tibetan language, we may imagine a language that is quite different from all current Sino-Tibetan languages. However, if we go back 2,000 years to Old Chinese, the Chinese language of that time might have been more similar to the original Sino-Tibetan language. So now let's turn to Chapter 3 to look at the sounds of Old Chinese; the grammar of Old Chinese will be taken up in Chapter 6 when we talk about Classical Chinese.