

1 Introduction

Chinese is the native language of the *Han* people, who form the largest ethnic group in China with over 90 percent of the total population. The Chinese language consists of seven mutually unintelligible dialect families, each of which contains many dialects and the largest of which is the Mandarin dialect family. In the broad sense, the word *Chinese* refers to all varieties of the language spoken by the *Han* people. In the narrow sense, Chinese or Mandarin is also used to mean Standard Chinese or Standard Mandarin, the official language of mainland China and Taiwan. Since there are major differences in the sound systems among the major dialect groups (cf. §12.1), this book will mainly focus on the sounds of Standard Chinese.

This introductory chapter has three goals. First, it provides basic background about the Chinese language in general and Standard Chinese in particular (§§1.1–1.4). Second, it presents a brief introduction to phonetics and phonology to set the foundation for the discussion of the subsequent chapters (§1.5). Third, it gives an overview of the topics and organization of the book.

1.1 The Chinese language family

The Chinese language family is genetically classified as a major branch of the Sino-Tibetan language family. The different varieties of Chinese can be grouped into seven dialect families, each of which consists of many dialects. The Mandarin dialects (or the northern dialects), spoken by more than 70 percent of Chinese speakers in the northern and southwest regions of China, can be further divided into four subfamilies: northern, northwestern, southwestern, and Lower Yangzi. The Beijing (or Peking) dialect, which forms the basis of Standard Chinese, is the best-known Mandarin dialect. The Wu dialects are spoken by more than 8 percent of Chinese speakers in the coastal area around Shanghai and Zhejiang Province. In Guangdong and Guangxi Provinces and

2 Introduction

in Hong Kong, the Yue dialects are spoken by 5 percent of Chinese speakers. Cantonese is a Yue dialect spoken in and around the city of Guangzhou (or Canton) and Hong Kong, as well as many traditional overseas Chinese communities. The speakers of each of the remaining four dialect families constitute less than 5 percent of the Chinese-speaking population. The Min dialects, consisting of northern Min and southern Min subfamilies, are spoken in Fujian Province, Taiwan, and some coastal areas of southern China. The Min dialect spoken in Taiwan, which is a variety of southern Min, is often called Taiwanese. The Hakka dialects are found near the borders of Guangdong, Fujian, and Jiangxi Provinces and widely scattered in other parts of China from Sichuan Province to Taiwan. Finally, the Xiang dialects are spoken in Hunan Province and the Gan dialects in Jiangxi Province.¹

The different varieties of Chinese are traditionally referred to as *regional dialects* (*fāngyán* ‘regional speech’) although the different dialect families and even some dialects within the same family are mutually unintelligible and could be considered different languages. For example, we can think of Mandarin and Cantonese as two different languages of the Chinese language family, just as Portuguese and Italian are two different languages of the Romance language family. In fact, some linguistics scholars prefer the term *Chinese languages* for those mutually unintelligible varieties. However, the tradition persists partly because all these varieties of Chinese share the same written language and a long tradition of political, economic, and cultural unity. For the moment, let us follow the tradition and refer to different varieties of Chinese as dialects and this issue will be discussed again in §12.1.

1.2 Standard Chinese

Standard Chinese (henceforth SC) is called *Pǔtōnghuà* ‘common speech’ in China, *Guóyǔ* ‘national language’ in Taiwan, and *Huáyǔ* ‘Chinese language’ in Singapore. Other English terms for SC include Standard Mandarin, Mandarin Chinese, or Mandarin. As the official language of China and Taiwan, SC is used in school and universities and serves all official functions. On national radio and television broadcasts, SC is the language used, but on regional stations, local dialects may be used in addition to SC.

In the early twentieth century, the standard pronunciation of SC was established and promoted by the Republic of China as *Guóyǔ* ‘national language.’ After 1949, when the People’s Republic of China was founded, SC was renamed as *Pǔtōnghuà* ‘common speech’ and defined as ‘the common language of China,

¹ This classification of Chinese dialects is based on Li and Thompson (1981:3) and Ramsey (1989:87). For a map of Chinese dialects, see Li and Thompson (1981:4), Ramsey (1989: Figure 6), Lyovin (1997: Map VIII), and the websites in Appendix C.

based on the northern dialects, with the Peking phonological system as its norm of pronunciation’ (Norman 1988: 135).² The lexical and grammatical expressions of SC are based more broadly on the northern Mandarin dialects but exclude specific local expressions including those used in the Beijing dialect. Although the pronunciation of SC is based on the phonology of the Beijing dialect, this does not mean the two have identical phonological and phonetic systems. For example, RHOTACIZED vowels (§8.2 and §3.4.5) are much more common in the Beijing dialect than in SC.

Although SC is taught in school and used in broadcasts, in reality the pronunciation of SC speakers is by no means uniform. What is considered to be the standard accent of SC tolerates a range of slightly different pronunciations. This phenomenon is common for any so-called standard language; for example, what is considered to be standard English in North America also covers a range of slightly different accents. In addition, there are different norms of SC in China, Taiwan, and Singapore, just as there are different norms of standard English in different English-speaking countries or regions. For example, in Taiwan and Singapore, the use of NEUTRAL TONE and rhotacized vowels is much less common than that in mainland China (see §12.2). The development of different norms of SC is mainly due to socio-political separation and the influence of the local dialects.

SC is generally associated with good education, authority, and formality, but educated people and government officials do not necessarily have the prescribed pronunciation of SC. Most Chinese learn to speak SC only after they have acquired their regional dialects and may learn from schoolteachers who do not have correct SC pronunciation themselves. In general, local dialects are used with family members and sometimes in public places, whereas SC is used more in schools and in workplaces (Chen 1999:54–5). In addition, many Chinese speakers regard SC simply as a practical tool of communication and often retain their local accents when speaking SC, especially within their local communities. Since there are so many different regional dialects, there are as many dialect-accented SCs or local SCs. More discussion of different varieties of SC will be given in §§12.2–12.3.

1.3 Tone, syllable, morpheme, and word

Chinese is a TONE language, a language in which changes in the PITCH of the voice can be used to denote differences in word meaning. We can think of tone as a third type of speech element in addition to consonants and vowels.

² For more details on the history of the establishment and promotion of standard spoken and written Chinese, see Chen (1999).

English makes use of consonants and vowels to form different words: *bad* and *pad* differ in one consonant and have different meanings; *bed* and *bad* have different vowels and also have different meanings. In addition to consonants and vowels, Chinese also uses tone to differentiate word meaning.

The examples in (1) from SC illustrate that in words with identical consonant and vowel combination, differences in tone are used to signal differences in meaning. The pitch value in the third column is based on a scale of 1 to 5, with 5 indicating the highest pitch and 1 the lowest (Chao 1930, 1968:26). In SC, the five levels of pitch distinguish four tones. For example, for the word ‘hemp’, the pitch starts in the middle of the pitch range (pitch level 3) and moves higher to pitch level 5. Traditionally, for ease of reference, the four tones are labeled as tone 1 to tone 4, as shown in the ‘tone number’ column (see also §4.2.1, example (7)). In the *pīnyīn* romanization system of SC (see §1.4 below), the tonal mark is placed on the vowel. It is important to note that, as a third type of speech element, tone is not an inherent feature of a vowel but can be viewed as a property of the whole SYLLABLE (see §4.1.3).

🔊 (1) Four tones in SC

C+V	TONE/PITCH PATTERN	PITCH VALUE	TONE NUMBER	PĪNYĪN	MEANING
ma	high level	55	tone 1	mā	‘mother’
ma	high rising	35	tone 2	má	‘hemp’
ma	low falling-rising	214	tone 3	mǎ	‘horse’
ma	high falling	51	tone 4	mà	‘to scold’

C = consonant V = vowel

In general, each Chinese syllable bears a tone. A syllable is a PROSODIC UNIT for carrying tone and STRESS. For example, in English, the word *system* has two syllables, *sys* and *tem*, with the first syllable as the STRESSED SYLLABLE. A Chinese word like *xuéxiào* ‘school’ has two syllables and two tones, a high rising tone (tone 2) on the first syllable *xué* and a high falling tone (tone 4) on the second syllable *xiào*. Chapters 4 and 9 will have further discussion of SC tone.

Chinese is typically classified as an analytic or isolating language in which each MORPHEME is usually a word. A morpheme is the smallest meaningful unit in a language. For example, in English, the word *uncontrollable* is formed by three morphemes: the word *control* is followed by the SUFFIX *able* and preceded by the PREFIX *un*. A free morpheme like *control* can stand alone as an independent word. On the other hand, an AFFIX (a prefix or a suffix)

is a bound morpheme that must be attached to a STEM (i.e. a morpheme to which an affix is added). AFFIXATION is the general process of adding an affix to a stem. Specifically, SUFFIXATION is the process of adding a suffix after a stem and PREFIXATION is the process of adding a prefix before a stem. Unlike English and many other languages, Chinese has very few prefixes and suffixes to form a complex word. On the other hand, modern Chinese has a great number of compound words similar to such compounds as *street light* and *wool sweater* in English. For example, *jiēdēng* ‘street light’ consists of two morphemes: *jiē* ‘street’ and *dēng* ‘light’, and *máoyī* ‘wool sweater’ can be decomposed to *máo* ‘hair, wool’ and *yī* ‘clothing’.³

Chinese is also often referred to as a MONOSYLLABIC language, which means that almost all words contain only one syllable. In English there are a great number of POLYSYLLABIC words. Whether or not Chinese words are monosyllabic depends on how a *word* is defined in Chinese, but to reach a consensus on such a definition has proved to be unexpectedly difficult in Chinese, partly because of the general lack of AFFIXATION in Chinese word formation.⁴ The term *zì* ‘character’ refers to a graph in the Chinese writing system (see §1.4 below) that corresponds to a morpheme and is one syllable in length. If each Chinese character is equivalent to a word, then Chinese words are indeed monosyllabic. However, if a word is defined as an independent basic unit for forming sentences, polysyllabic forms such as *xuéxiào* ‘school’ and *rúguǒ* ‘if’ should be considered single words although they consist of two syllables and they are written with two characters. The characterization of Chinese as being monosyllabic fits much better with classical Chinese where over 90 percent of words are monosyllabic. In modern Chinese, however, 95 percent of morphemes are monosyllabic,⁵ but about half or more than half of the words are polysyllabic and consist of more than one morpheme (cf. Chen 1999:138–9).

To summarize, in modern Chinese, each syllable generally bears a tone, most morphemes are monosyllabic, and words may consist of one or more morphemes and hence may be monosyllabic or polysyllabic.

1.4 Chinese characters, romanization, and pronunciation

Chinese has a logographic writing system in which each character represents a morpheme, whereas an alphabetic writing system as used in English employs

³ In fact, *yī* is not usually used as a single word; it is combined with *fú* ‘clothes’ to yield *yīfu* ‘clothes’. Some non-affix morphemes in modern Chinese have to be combined with other morphemes to form independent words. For more information about Chinese morphemes and words, see Li and Thompson (1981: chapters 2 and 3) and Packard (2000).
⁴ See Duanmu (2000: chapter 5) and Packard (2000) for more details.
⁵ Polysyllabic morphemes include words such as *húdié* ‘butterfly’ and most transliterated loanwords, such as *nīùyǔè* ‘New York’.

a character (or letter) or combination of characters to represent the speech sounds. As mentioned above, each Chinese character is one syllable in length and the majority of Chinese morphemes are monosyllabic. Since a morpheme is the smallest meaningful linguistic unit, each Chinese character expresses some meaning. This does not mean there is no way to get a hint of pronunciation from the characters. In fact, over 90 percent of Chinese characters consist of one subcomponent denoting meaning and another denoting the pronunciation (Chen 1999:141). In (2a–c), all three characters are pronounced with the same consonant and vowel combination.

(2)

	CHARACTER	C+V	TONE	MEANING	PĪNYĪN
a.	马	ma	214	‘horse’	<i>mǎ</i>
b.	妈	ma	55	‘mother’	<i>mā</i>
c.	蚂	ma	214	‘ant’	<i>mǎ</i>
d.	女			‘female’	<i>nǚ</i>
e.	虫			‘insect’	<i>chóng</i>

The character in (2a) is used in (2bc) to denote the pronunciation of the consonant and the vowel but not the meaning. That is, all three characters in (2a–c) share the same component 𠂔 and are pronounced with the same consonant and vowel combination [ma], but the meaning of ‘horse’ for 𠂔 has nothing to do with ‘mother’ (2b) or ‘ant’ (2c). The subcomponent at the left side of (2b) means ‘female’ and that of (2c) means ‘insect’, as illustrated in (2de). The left-side subcomponents in (2bc) do not denote possible pronunciation but do contribute to the meaning: ‘mother’ is female and ‘ant’ is a kind of insect. Although most Chinese characters contain a subcomponent to signal possible pronunciation, there is still a relatively high degree of arbitrariness between a written character and its actual pronunciation.⁶

The Chinese writing system as a well-developed system was established roughly in the fourteenth century BC and since then has undergone several major stages of development. The total number of Chinese characters is now around 56,000 but the number of most common characters are 2,500 and a college graduate is expected to recognize at least 3,500 characters (Chen 1999:136). The high number and complexity of Chinese characters and the difficulty of

⁶ However, see DeFrancis (1984) for a different view and for more discussion of the nature of Chinese characters. See also Chen (1999: chapter 8) for the basic features and history of the Chinese writing system.

learning them were often regarded as obstacles to achieving a high literacy rate and modernization. Simplification of the writing system thus became an issue during the first half of the twentieth century. In 1956 the People’s Republic of China promulgated the Scheme of Simplified Chinese Characters and since then this new set of simplified characters has been used in China. However, the traditional characters are still used in Taiwan and many traditional overseas Chinese communities. Throughout this book, when Chinese characters are provided in examples, the simplified characters are adopted.

The romanization systems designed to indicate the pronunciation of Chinese characters were first developed by Western missionaries in China. The first phonographic writing of Chinese promoted by the government was *zhùyīn zìmǔ* ‘sound denoting letters’ or *zhùyīn fúhào* ‘sound denoting symbols’, which was used as a tool to teach and annotate the pronunciation of characters before 1958 in China and has been in continuous use in Taiwan. In this system, the roman alphabet is not adopted and instead a set of simple characters is used; for example, **ㄇ** represents the sound [m] and **ㄞ** the vowel sequence [ai]. In contrast, *hànyǔ pīnyīn* ‘Chinese sound spelling’ or simply *pīnyīn* ‘sound spelling’, which replaced *zhùyīn zìmǔ* after 1958 in the People’s Republic of China, adopts the roman alphabet. The *pīnyīn* system, which has become the standard transcription system of Chinese words, has been taught in school in China, is the most popular romanization system taught in school outside China, and is used as the input system in Chinese word processing on computers. Before *pīnyīn* became commonly in use, the Wade-Giles romanization system, which was created by Sir Thomas Wade and modified by Herbert A. Giles in his *Chinese–English dictionary*, published in 1912, served as the standard transcription system in scholarly works in English (Norman 1988:173). A less commonly used system is the Yale system, which was developed from the *Dictionary of Spoken Chinese* issued by the War Department in the United States in 1945 (Norman 1988:174–5). In Taiwan there is an official romanization scheme for the transliteration of Chinese proper names, which is similar but not identical to *pīnyīn*; however, many people continue to use the Wade-Giles system or base the transliteration on the pronunciation of the local dialects.⁷ In this book, we use only the *pīnyīn* system.

However, the *pīnyīn* system is not really a phonetic transcription system. In fact, no romanization or alphabetic systems provide a perfect match between an alphabetic letter and the actual pronunciation. For example, in English, the

⁷ For more details on the system of *zhùyīn fúhào* and the official romanization system in Taiwan, see Chen (1999:181 and 190). For a comparison of the Wade-Giles system, the Yale system, and *pīnyīn*, see the Internet resources in Appendix C.

8 Introduction

letter *i* is pronounced differently in *live* and *life*, and in *pīnyīn* the letter *i* is also pronounced differently in *sī* ‘silk’ and *jīa* ‘home’. To accurately transcribe pronunciation, we have to use a phonetic transcription system such as the INTERNATIONAL PHONETIC ALPHABET (IPA). The IPA is the standard transcription system used by linguists to represent the sounds of all human languages. Since in this system each attested human language sound is represented by a unique phonetic symbol, it becomes possible to transcribe and describe phonetically different sounds that are not differentiated by orthography or romanization systems. In chapter 6, we will see a systematic comparison between *pīnyīn* and the IPA transcriptions of SC sounds.

1.5 Phonetics and phonology

Linguistics is the scientific study of human language that investigates: (i) what the structure of language is and in what aspects languages are similar and different; (ii) how language is acquired and processed and how language works in the human cognitive system (psycholinguistics); (iii) how the brain functions in language production, perception, and processing (neurolinguistics); (iv) how language is used in different societies and speech contexts (sociolinguistics); (v) how language changes over time (historical linguistics); and (vi) how the knowledge derived from linguistic studies is applied to other areas such as language teaching, speech disorders, and computer science (applied linguistics). Different aspects of language structure are studied by different subfields of linguistics. PHONETICS studies speech sounds: how they are produced and classified, what their physical properties are, and how they are perceived. PHONOLOGY examines the sound system of language: how speech sounds are organized to form a system for encoding linguistic information. MORPHOLOGY is the study of word formation: how words are constructed out of smaller meaningful units (i.e. morphemes); SYNTAX is the study of sentence structure; semantics is the study of meaning; and pragmatics the study of meaning in context, i.e. how the meaning and interpretation of a word or sentence depends on the context in which it is used. The linguistic grammar, then, is the combination of all these different aspects of language structure. It is important to note that a linguistic grammar *describes* the language that speakers actually use, and, unlike traditional grammar books, does not *prescribe* what a correct grammar should be. To know more about what linguistics is, see the suggested introductory linguistics textbooks given in Further Reading.

In this book we study the phonetics and phonology of SC. The following subsections provide a brief introduction to phonetics and phonology to set the stage for the discussion and more advanced topics in the remaining chapters.

1.5.1 Phonetics

Phonetics consists of three areas of inquiry. **ARTICULATORY PHONETICS** describes how speech sounds are produced and how sounds are classified according to their articulatory properties. **ACOUSTIC PHONETICS** examines the physical properties of speech sounds such as duration, frequency, and intensity; **PERCEPTUAL PHONETICS** (or auditory phonetics) is the study of the perception of speech sounds. In the discussion of SC phonetics, this book focuses on articulatory phonetics with supplementary information from acoustic and perceptual phonetics.

Speech sounds are produced by modifying the airstream. Most speech sounds are made when the air from the lungs is pushed through the larynx and the oral and nasal cavities. Sounds created in this way are said to use the pulmonic egressive airstream mechanism.⁸ Different speech sounds are produced by modifying this airstream at different points along the pathway of the airflow (see Figure 2.2 in §2.1.2).

The air coming from the lungs may be modified at the larynx (sometimes called the voicebox), which is located at the top of the trachea (or windpipe) and houses the vocal folds. The front of the larynx (Adam’s apple) protrudes slightly at the front of the throat. This is the first point where the flow of the airstream can be modified. The **VOCAL FOLDS** are folds of muscle that can close together or move apart; the opening between the vocal folds is called the **GLOTTIS**. When the vocal folds are held close together and made to vibrate by the air pushed repeatedly through the vocal folds, a **VOICED** sound is produced; on the other hand, when the vocal folds are open and the air flows through the glottis freely, a **VOICELESS** sound is produced. For example, in English, [z] in *zip* is a voiced sound and [s] in *sip* is a voiceless sound. To help detect if a sound is voiced or voiceless, put your fingers lightly on your throat (close to where the Adam’s apple is) and say [z] for a few seconds and [s] for a few seconds and then say [zzzzsssszzzzssss]. You should be able to feel vibration inside the larynx when you produce [z] but no such vibration when you say [s]. You can use the same method to find out what other sounds in English are voiced and voiceless. For example, [v] in *vine*, [m] in *man*, [l] in *life*, and all vowels are voiced sounds; the first sounds in *five*, *she*, and *thing* are voiceless sounds.⁹

Another important component of human physiology for speech production is the **VOCAL TRACT** above the larynx (supralaryngeal vocal tract), which

⁸ Other airstream mechanisms are possible for some less common sounds (Ladefoged 2001: chapter 6). SC and English, like most languages, use only the pulmonic airstream mechanism.
⁹ Note that this voicing test does not work well for the voiced–voiceless pairs of sounds such as [b]–[p], [d]–[t], and [g]–[k]. See §2.1.1.

includes the pharynx (the passage connecting the larynx and the oral cavity in the mouth), the oral cavity in the mouth and the nasal cavity within the nose. The flow of air can be modified at various locations in the vocal tract to produce different sounds. For example, to produce a [b] sound as in *bay*, we obstruct the airstream temporarily by closing the lips, and to produce a [d] sound in *day*, we obstruct the airstream by placing the tongue tip or tongue blade (the frontmost part of the tongue) at the back of the upper teeth or the ALVEOLAR RIDGE (the protruding bony area behind the upper teeth). The different points at which obstruction can be made are the PLACES OF ARTICULATION. The flow of air can also be modified with different degrees of obstruction. The different ways in which a sound is modified are the different MANNERS OF ARTICULATION. Compared to consonants, vowels have relatively free flow of air through the vocal tract. Among consonants, a sound like [d] makes a complete obstruction of the airstream whereas a sound like [z], which has the same place of articulation as [d], has a narrow opening between the raised tongue and the alveolar ridge to let the air squeeze through the narrow channel. A consonant like [d] is called a STOP because the airstream is completely obstructed and a consonant like [z] is called a FRICATIVE because the air pushed through the narrow channel produces friction noise. In chapters 3 and 4, we discuss how consonants and vowels are made and classified according to the status of the vocal folds, the place of articulation, and the manner of articulation.

When a sound is produced, the flow of air is converted to sound waves that can be transmitted through the air for the listener to perceive. The vibration rate of the vocal folds determines the FUNDAMENTAL FREQUENCY of a sound wave. If the vocal folds complete each cycle of vibration 100 times in a second, then the fundamental frequency is 100 Hz. A high tone has higher pitch and higher frequency and a low tone has lower pitch and lower frequency. A rising tone then has a pitch pattern of change from a lower pitch to a higher pitch. The phonetic properties of different tones in SC is discussed in chapter 4.

Each speech sound has its own set of acoustic properties so that the listener can distinguish one sound from another. The various modifications of the airstream in speech production create different patterns of sound waves. That is, the source of a sound wave produced by the lungs and the vocal folds is modified in different ways in the vocal tract to yield different acoustic characteristics (different patterns of sound waves) for different sounds. The human speech production system is similar to a musical instrument: a musical instrument also has a source of sound (e.g. the air blown across a flute’s mouth hole) and produces different musical notes by modifying the sound in different ways in