1

Some Preliminaries

1.1 Readership

This book is written for the English-speaking audience in the world who cannot read publications in Chinese, especially those who cannot read texts in Classical Chinese. We have selected the essentials of our own studies from the past twenty years or so to provide a comprehensive picture of the evolution of Chinese grammar over the past three millennia. Thus we have made technical adjustments to increase the readability of the contents, including the following four features. First, traditional Chinese characters and the corresponding Pinyin transcriptions are provided only in the numbered examples; no Chinese characters are provided for Chinese expressions in the narrative parts. Second, for the historical examples and the examples in the northern dialects, only Pinyin forms representing the phonological system of standard Mandarin are provided, though there are apparently phonological variations across dialects. Only when the phonetic forms of the examples are relevant to the analysis are International Phonetic Alphabet (IPA) representations provided. Examples from other major dialect families are given in the IPA form according to the phonological systems of the representative subdialects. Third, only the general titles of the ancient texts from which examples are cited are indicated; the subtitles of these texts are omitted because they are meaningless to Western readers. Fourth, we do not adopt the common practice of dating in Chinese historical linguistics by using the dynasty, such as the Han Dynasty and the Tang Dynasty, as the temporal frame to describe the historical development. Instead, the Western calendar is adopted to date examples and the development of the language.

All analyses are made within contemporary linguistic frameworks, mainly grammaticalization theory, topological linguistics, cognitive linguistics, construction grammar, functionalist linguistics, and general linguistics. Our work is neither a simple description of the changes nor an exhaustive list of all changes in the history of Chinese. This book is not written encyclopedically; only parts that are theoretically significant will be discussed in detail. However, all major changes in the history of Chinese are covered in this book.

1.2   **A Natural Laboratory**

The Chinese language has a well-documented history of approximately 3,300 years that is actually invaluable for studying the languages and civilization of all human beings. Although there are roughly 6,700 different languages in the world, Chinese possesses not only the longest history but also the greatest amount of documentation that accurately and completely records detailed developments at each stage. In this regard, only Greek may be comparable with Chinese. To understand the scientific value of historical documents of the Chinese language, let us consider some parallel examples in the natural sciences. There are two types of law in nature: one type, such as the law of free fall, needs to be proven in a human-designed laboratory; the other type, such as Einstein's relativity theory, can be tested only in a natural laboratory, namely the natural world.

First, we consider how to identify and prove the first type of natural law. Aristotle (384–322 BC), an ancient Greek philosopher, discussed the law of falling objects in *Physics* (Book VII), and from then to the sixteenth century AD it was generally assumed that the speed of a falling body was proportional to its weight. That is, a ten-kilogram object was expected to fall ten times faster than an otherwise identical one-kilogram object through the same medium. This "intuitively correct" law was found to be wrong by physicists in the sixteenth century AD, when Galileo (1589–1592) is said to have first discovered the law by dropping two objects of unequal mass from the Leaning Tower of Pisa. However, the law of free fall can be confirmed as a scientific law that is entirely accurate only in human-created ideal conditions of a uniform gravitational field without air resistance. Likewise, identifying the structure of atoms and capturing ions require sophisticated equipment such as electronic microscopes and accelerators that are designed by scientists.

Next, we would like to provide an example of a natural law that can be tested only in nature. On the basis of his new theory of general relativity, which was proposed in 1911, Einstein calculated that light from another star should be bent by the sun's gravity. In 1919, that prediction was confirmed by Sir Arthur Eddington during the solar eclipse on May 29. Additionally, the theory of general relativity proposed by Einstein approximately 100 years ago posited that gravitational waves exist in the universe. To prove this hypothesis, the American government spent a huge amount of money building two laboratories in northern and southern areas. In 2016, scientists in the two laboratories announced that they had made the first direct observation of gravitational waves originating from the merging of a binary black hole system.

Likewise, linguistic rules also fall into two types: one type is rules that can be proven by psychological experiments or statistical surveys of contemporary language; the other type is rules that can be discovered and tested only in the historical development of

a language. For example, how construal influences grammatical forms may be identified by a psychological experiment, but how construal affects the recruitment of lexical sources for a grammatical category, such as the passive, requires a comprehensive investigation of several millennia of historical data. We can benefit greatly from the history of a language if we view it as a natural laboratory for testing the validity of any theoretical hypotheses. According to our own research experiences, the evolution of a language is especially invaluable for exploring the following linguistic issues.

(a) Exploring the relationship between cognition and language. One of the unresolved yet most challenging questions is why all human languages must constantly change over time and no language can remain the same forever. Answering this question involves interdisciplinary collaboration among linguistics, psychology, neurology, and cognitive science, but any hypothesis can be tested only in the evolution of languages.

(b) Testing theoretical hypotheses of different linguistic frameworks. For example, the usage-based model assumes that repetition and frequency are mainly responsible for the innovation of grammatical forms (Bybee 2006: 269–278). Similarly, cognitive linguistics claims that grammatical structures develop out of the entrenchment and enforcement of certain linguistic expressions (Langacker 1987: 59–60). Additionally, grammaticalization theory suggests that pragmatic inference is a major factor in triggering grammatical change (Hopper and Traugott 2003: 71–98). In contrast, generative linguistics believes that language consists of principles and parameters and that the former are immune to change but the latter may change due to discrepancies in language acquisition (Lightfoot 2013). In our view, however, the validity of these hypotheses must ultimately be tested using diachronic evidence.

(c) Discovering universal correlations across languages. For example, Greenberg (1966a) identified a universal correlation: if a language takes SVO as its basic word order, its relative clause follows the head noun. After more than 1,000 languages were investigated, the results indicated that Chinese was the sole counterexample to this universal correlation (Dryer 1992, 2007). However, if we examine the diachronic data of Chinese, it is clear that Chinese was consistent with the correlation. In addition, the motivation for causing the change from consistency to inconsistency reveals an essential property of the interaction among different constructions.

(d) Proving concrete argumentation about the relation between different grammatical categories. For example, it has been generally accepted that, historically, the subject grammaticalized out of the topic. Logically, this

hypothesis entails that English grammar be more developed than Chinese grammar because the former is subject-prominent and the latter topic-prominent (Li and Thompson 1976, Hopper and Traugott 2003: 28). A simple comparison between diachronic and synchronic data clarifies whether this view can be true.

(e)   Revealing the relationship between different constructions. For example, it is assumed that the passive structure is transformed from an underlying active form, a key point that Chomsky (1957) established in his generative linguistics. Since then, the correlation between the passive and the active has occupied the central position in creating other theoretical frameworks, such as cognitive linguistics, lexical functional grammar, and relation grammar. In our view, the real relationship between the passive and the active can be seen only from diachronic change in the grammar.

The earliest written records of Chinese were the oracle bone inscriptions, which can be dated back to the thirteenth century BC. Since then, Chinese has been well documented in numerous genres, such as poetry, dialogues, novels, books of history, philosophy, politics, and many other subjects that objectively reflect the situation of the spoken language at each stage. Although Chinese is not the earliest language in the world to have been recorded by a writing system, the other earlier languages all ceased to be used at some point in history. For example, the oldest known written language is Sumerian, which dates back to at least the thirty-fifth century BC. The earliest proof that the written Sumerian language existed was the Kish Tablet, which was found in Iraq. Sumerian is older than Egyptian, but it lasted as a spoken language only until the twentieth century BC, when it was replaced by another language, called Akkadian.

The longer the history of a language is, the better chance we have of identifying the regularity of the development of the grammar. A grammaticalization process often takes several hundred years, and may take over a millennium. An extreme case is the development of reduplication, which took nearly 3,000 years to reach the completion point (for details, see Section 10.2). Specifically, reduplicated forms of adjectives and adverbs were already commonly used in texts composed around the eleventh century BC, but the reduplicated forms of nouns first appeared in texts around the third century AD, those of classifiers emerged around the eighth century AD, and not until the fourteenth century AD did verb reduplications start to occur. Why the reduplicated forms of different word classes emerged at such different periods is the key factor in understanding the grammatical properties of these forms and the mechanism of language changes.

As we know, the writing system of Chinese is not phonetic in nature. It is also by and large not pictographs, which account for merely a small portion of Chinese characters; approximately 80 percent are pictophonetic, with one part of a character indicating the semantic category and the other part roughly identifying the sound. As the phonological system developed, the writing system remained highly stable. Thus this stability enables us to investigate the whole history of the language without much difficulty. The original texts of poems and prose works have been adopted in textbooks of Chinese language teaching even in primary schools, and the greater proportion of Chinese textbooks are those of Classical Chinese. People who have obtained a BA in Chinese language and literature are generally able to understand texts composed in ancient times. This reading ability is crucial for an individual researcher to discover the regularity of the evolution of the Chinese language. Therefore numerous scholars in the circle of Chinese linguistics have studied Classical Chinese, and a tremendous number of books and papers have been published in this area. These collective efforts are very helpful in exploring the development of the language. By comparison, the earliest English texts were composed around the seventh century AD, and the writing system was basically phonetic, although it has changed over time. Only limited scholars who have undergone special training can understand the ancient texts, which still does not guarantee that they will understand texts composed at a different stage. As Barber (1995: 100) pointed out, people today are almost entirely unable to read texts in English that were composed around the twelfth century AD. In contrast, most people who graduate from high school in China find it comfortable and enjoyable to read the poetry, prose, and other literary genres composed after the tenth century AD.

Another advantage of investigating the history of Chinese is that the language system has never been fundamentally influenced by other languages due to low language contact, and its vocabulary, phonology, and grammar have been developing on their own path, without disruption by any foreign language contact. In contrast, almost every language of the Indo-European family, such as English, French, German, and Italian, has undergone fundamental changes due to language contact. Thus their development paths were often disrupted and altered by other languages, which interfered with and blurred the regularity of their development.

In short, the long, undisrupted, and well-documented history of the Chinese language makes it the best natural laboratory to experiment with scientific problems in relation to human language.

1.3   Periodization of the Chinese Language and the Dating of Texts

There is no generally accepted periodization of the history of the Chinese language in the literature, and various labels are used, such as Archaic Chinese, Old Chinese,

Classical Chinese, Medieval Chinese, and Modern Chinese, which often refer to quite different time spans. There are several periodizations of the Chinese language, which are based on different criteria, such as syntax, phonology, or lexicon (Wang 1980, Lü 1985, Peyraube 2017). Since this book is targeted at readers in the English-speaking world, we adopt the terminology used for describing the stages of the development of English that is widely accepted in the English literature (e.g. Hopper and Traugott 2003: xx, Kytö and Päivi 2016: 1–18).[1] The primary purpose of dividing the stages of the history of a language is to facilitate the description of its development. According to our research experiences, therefore, the following periodization is most convenient for describing the historical development of the Chinese language.

The periodization is further refined by adding either an "early" or a "late" modifier. When these modifiers are added, the term refers to the first or last three centuries of the period; for example, "Early Medieval Chinese" means a period from the second century BC to the first century AD. When periodizing the history of the Chinese language, we have taken into consideration the periodizations of Wang (1980: 35) and Lü (1985: 1), both of which are highly influential in the field.

Table 1.1 *The periodization of the Chinese language*

Old Chinese 1300 BC–200 BC
    Early Old Chinese 1300 BC–800 BC
    Middle Old Chinese 800 BC–400 BC
    Late Old Chinese 400 BC–200 BC
Medieval Chinese 200 BC–AD 1000
    Early Medieval Chinese 200 BC–AD 300
    Middle Medieval Chinese AD 300–AD 700
    Late Medieval Chinese AD 700–AD 1000
Modern Chinese AD 1000–1900
    Early Modern Chinese AD 1000–1300
    Middle Modern Chinese AD 1300–1700
    Late Modern Chinese AD 1700–1900
Contemporary Chinese AD 1900–present

---

[1] The stages of English are listed as follows:

Old English	AD 600–1125
Middle English	AD 1125–1500
Early Modern English	AD 1500–1750
Modern English	AD 1750–1950
Contemporary English	AD 1950–present

The above four stages are taken into account the timing of the establishments of three major dynasties. Specifically, the boundary between Old and Medieval Chinese is the approximate time when the Qin dynasty was established in 221 BC and the Han dynasty in 206 BC, which signals the beginning of the feudal society in Chinese history. The division between Medieval Chinese and Modern Chinese refers to the transition period from the Tang dynasty to the Song dynasty: the former ended in AD 907, and the latter was established in AD 960. The division between Modern Chinese and Contemporary Chinese represents the end of the Qing dynasty (precisely AD 1912) and the beginning of the new era of the Republic of China (precisely AD 1912). Therefore our periodization is compatible with the temporal frames that a vast majority of scholars have become accustomed to using in their research. By noting these correlations, we do not mean that the development of the grammar has any relation to the rise and fall of different dynasties. The evolution of language always follows its own paths, which are determined mainly by its internal structural factors and the collective cognition of the language community, and changes do not happen at the same pace. As far as the same duration of a period is concerned, in some periods the language was quite stable and no dramatic changes happened, but in other periods many dramatic changes happened, possibly even affecting the overall texture of the language.

Theoretically, any change can be dated to a particular time, date, or even minute. However, it is impossible and unnecessary to do so in a historical investigation. Although any language is always in the process of evolution, no change happens in a single day. Change is always gradual, and the language system typically remains stable over quite a long period of time. This book adopts fifty years as a basic time unit in discussing historical development. The dates that are provided after each example are understood to represent the linguistic system around that time rather than specifying the precise time when the text was actually composed. Our way to date the examples is to identify the midpoint of the life span of the author and see whether it is closer to the upper or lower time unit boundary of a fifty-year period. For example, the life span of Confucius lasted from 551 to 479 BC, and the midpoint is 514 BC; hence examples from the *Analects* are dated as 500 BC. In many cases, there is not enough information to identify both the authors and the time of composition; in those cases, we can provide only a very rough dating. For example, the *Classic of Poems* is a collection of folk songs that were composed from the eleventh to the seventh century BC, so all we can say is that it basically reflects the language system of Early Old Chinese.

1.4 **Historical Texts**

To explore the evolution of the grammatical system, the ideal situation is that there are always enough vernacular texts available for each stage of language development to

accurately and entirely record the spoken language at each particular period of time. However, this kind of text is rare, and most historical texts are actually a mixture of spoken and written and are even infused with dialectal expressions. We deal with these complex situations in the following ways. First, we carefully chose texts of the genres that were most likely to reflect the system of the spoken language at the time, such as folk songs, dialogues, drama scripts, and religious speeches for ordinary people. Second, since we are interested mainly in regular phenomena that had far-reaching effects on the development of the language, we tend to ignore usages that existed only in limited texts with short life spans, some of which might belong to other dialects. More importantly, we view the whole history of Chinese as an integral system when doing our research, and we try to put every phenomenon in its proper position in the overall historical context. In this case, we can easily detect which phenomena were meaningful for the development of the grammatical system.

To avoid our being biased by our own theoretical backgrounds and misled by irregular usages, the scope of the investigation must be wide enough to cover all the major texts of each period. We have read most of the representative texts that reflect the grammatical system at different times, which has given us a feeling for the general situation of the system. Additionally, electronic databases of ancient texts have enabled us to conduct a comprehensive statistical survey on any usage, which is crucial in two respects: first, many rules and principles can be revealed in the statistical data; second, these statistics can be used to test many theoretical hypotheses. Over the past two decades or so, we spent a tremendous amount of time reading the original texts, collecting a large number of examples, and manually conducting many statistical surveys. In addition, we have made use of three main important databases of ancient Chinese texts: (a) the Corpus of Ancient Chinese Texts at Academia Sinica of Taiwan, (b) the Chinese Corpus of Peking University, and (c) the Chinese Language Corpus of Beijing Language University.

According to our own research experiences, the most useful theoretical frameworks investigating the development of a language are grammaticalization theory, linguistic typology, cognitive linguistics, and construction grammar. However, our general philosophy or methodology is that we learn linguistics from language and our main goal is to discover rules, principles, and any other regular phenomena in the natural language. Thus any ideas and methods can be used in our analyses if they are helpful in realizing our research goals.

1.5   **The Structure of the Contents**

There are many ways to organize the contents of a book in historical linguistics, which are determined largely by the purposes of the research. For example, *The Cambridge History of the English Language* (Hogg et al. 2001), which consists of six large volumes,

is structured by dividing the language into four periods and two large areas, with one volume dedicated to each. Each of the six volumes contains roughly the same topics, such as phonology, semantics, and syntax, written by different scholars. And different scholars were invited to write about the same topic relative to different periods; for example, the author of the section on syntax during the period from the beginning to AD 1066 is Elizabeth Traugott, while the author on the same topic in the succeeding period from AD 1066 to 1476 is Olga Fischer. Structuring a book in this way is convenient for descriptive historical linguistics, but it is obviously disadvantageous for exploring the motivation, mechanism, and regularity of language development. One reason is that a complete grammaticalization process usually takes several hundred years, and some might even take more than a millennium to reach the completion point of their development. For instance, an extreme case in the history of Chinese is the development of reduplication, which took nearly three millennia to expand from adjectives to verbs. More importantly, the starting and end points of a developmental process will vary from one grammatical category to another.

Since the focus of this book is on the motivation, mechanism, and regularity of the evolution of Chinese grammar, mainly exploring the motivations and mechanisms behind language developments, we adopt a unique structure for the contents of this book, which distinguishes it from any other historical linguistics book, regardless of whether it is written in Chinese or any other language. Specifically, we structure and order the contents according to the natural relationship among the grammatical forms, and every change within the same cluster holds some cause–effect relationship. For example, the establishment of the resultative construction motivated the emergence of the disposal construction, the verb-copying construction, the aspect system, and verb reduplication; therefore these topics are arranged in the logical order. That is, they are discussed together within either one section or over several consecutive sections.

2

Copular Word and Construction

2.1 **Introduction**

The copular construction is one of the basic grammatical apparatuses required by daily communication; hence any language or a language at different periods of time always has some kind of grammatical device to perform the corresponding function.[1] Thus we begin with a fundamental change in copula structures from Old to Medieval Chinese, which in turn caused a series of consequent developments that significantly altered the texture of Chinese grammar at the time. The earliest documents show that the basic word order in Old Chinese was SVO, although there were some OV variations with special pragmatic values (for details, see Wang 1989: 198–216, Yang and He 2001: 784–813). In comparison, a typical SVO language, English, has a copula (i.e. *be*) that occurs between the subject and the complement, as ordinary verbs do, which shows personal agreement with the subject and inflects for tense and aspect. However, Old Chinese lacks such a copula verb, as there is no linking verb between the subject and the complement, and the copula construction at the time was obligatorily marked by a sentence-final particle *yě* that followed the complement, unlike other verbs. The copula construction in Old Chinese can be formulated as in (1).

(1)   Subj (*zhě*), NP *yě*.

The word *zhě* in parentheses is a demonstrative, optionally used as an anaphor to refer to the subject; there is a pause between the subject and the complement that is indicated by a comma in written texts, as illustrated below.

(2)   弓矢者，器也。(易經・繫辭下)
       Gōng   shǐ   zhě,   qì   yě.
       bow   arrow   ANAP   instrument   COP
       "Bows and arrows are instruments."

(*Yi Jing*, *Xi Ci*, 800 BC)

---

[1] In contrast, some other constructions may not be essential and they did not exist all the time; some might have disappeared from history, such as the verb co-ordinate construction (for details, see Chapter 4), and some might have been innovated at a certain stage, such as the disposal construction (for details, see Chapter 9).

10