Aspects of Classical Chinese grammar

Jinyuan Wu

April 26, 2025

Chapter 1

Introduction

1.1 The name of the language

This note is about Classical Chinese, the high variety of more than two millennia of diglossia in China. The language is known natively (in Mandarin Chinese) as 文言 ('lit. cultured speech') or sometimes 古文 ('lit. ancient articles') or 古汉语 ('lit. ancient Chinese'). Note that there were several stages of the development of Chinese and Classical Chinese is mostly (but not completely) based on Old Chinese (§ 1.2).

The language is sometimes known as *Wen-li* by Western missionaries, especially in Bible translation. This seems to be a misunderstanding of the word 文理, which is a nominal compound and means rhetorics (i.e. 文) and meanings (i.e. 理) of literature works. An educated person therefore would be described as "通文理" ('fluent in rhetorics and meanings'). Such a person of course would have decent understanding of Classical Chinese, and hence 文理 was probably mistranslated as "Classical Chinese", although the word was not natively used to refer to the latter.

The word 文言 is sometimes translated to *Literary Chinese* in English. We need to note that this translation can be misleading. Some Chinese written materials were definitely *not* written in 文言 i.e. Classical Chinese, like novels like *Journey to the West* or *Dream of the Red Chamber*: they were written in earlier forms of Mandarin. Buddhist scriptures translated to Chinese are more or less Classical but are also definitely influenced by the grammar of Middle Chinese. The non-Classical written materials, despite being important both culturally and linguistically, are beyond the scope of this work.

1.2 Historical background

Since there was no attempt at explicit and systematic grammatical standardization (§ 1.4.1), the prescriptive authority of Classical Chinese was a collection of canonical literature works consensually regarded as classical (§ 1.5). The whole canon was finished before the collapse of Han and therefore falls under the term Old Chinese. Both temporal and regional variances can be observed in Old Chinese texts, though, and not all varieties contribute to Classical Chinese equally. In this section, we briefly overview the history of Sinitic languages and analyze how they shape Classical Chinese.

1.2.1 The prehistory of the Sinitic family

1.2.1.1 The Trans-Himalayan family

The Sinitic language family, often collectively referred to as Chinese, is a branch in the Sino-Tibetan language family, which is also known as the Trans-Himalayan language family. While the internal structure of the Trans-Himalayan language family and consequently a reconstruction of Proto-Sino-Tibetan with broad consensus (at the level of e.g. Proto-Indo-European) remain to be explored, the shared origin of Chinese, Tibetan, Burmese, and multiple other language families however is beyond a reasonable doubt: for example, a list of cognates and a tentative Trans-Himalayan reconstruction can be found in Hill (2019).

Having established a "flat" relation between Chinese and other languages in the Trans-Himalayan area, we may be interested in the position of Chinese within Trans-Himalayan, and in particular, when it diverged from the rest of the language family. Note that this is to be decided by comparison of cognates, and *not* by typological features (§ 1.2.1.2). This is inevitably related to the internal structure of the language family. A consensual phylogeny based on the comparative method is still lacking. Some scholars take a rather pessimistic view and suggest that it is probably a better idea to remain agnostic about higher-order subgrouping and adopt a *Fallen leaves model*, i.e. a thorough enumeration of all genetically certain subgroups without any clear genetic relations stipulated above them.

Two Bayesian phylogenetic studies (Sagart et al. 2019; Zhang et al. 2019) both consider the Sinitic branch to be the earliest branch that diverged from the rest of the language family, thus supporting the traditional binary division between the Sinitic branch and the Tibeto-Burmese branch.[1] They however cannot replace human readable and interpretable comparative analyses, and currently, the name *Tibeto-Burmese*, strictly speaking, is still not guaranteed to be a node in the phylogenetic tree of the Trans-Himalayan family, but rather a catch-all term for everything non-Sinitic.

1.2.1.2 On "Sino-Tibetan typology"

Purported typological similarities and counterexamples Besides cognates, linguists occasionally talk about "Sino-Tibetan typology". Typological features often attributed to Sino-Tibetan languages include the isolating grammar, tonality, and monosyllabic roots.[2] These criteria was even once used to demarcate the boundary of the Sino-Tibetan language family, which later proved to be not reasonable. A refutation on the phonological part can be found in Matisoff (1973): features like tones and monosyllabic roots are not diachronically stable at all. As for morphological complexity, just think about how drastically the English inflectional morphology has simplified. Recent linguistic fieldwork has also revealed that several previously underdocumented Trans-Himalayan languages completely violate these generalizations. The Gyalrongic

[1] It should be noted that the huge divergence of contemporary Chinese languages from their Tibeto-Burmese cousins cannot be just explained by Neogrammarian sound change laws: language contact is an important factor (§ 1.2.1.5).

[2] Modern Sinitic languages are by no means monosyllabic, while Classical Chinese appears to be monosyllabic: and disyllable words in later stages of Chinese seem to be due to either borrowing or syntactic fossilization of previous analyzable phrases. We are going to cover some of these in Ch. 11. On the other hand, pre-Classical Old Chinese is possibly not strictly monosyllabic according to recent reconstructions (§ 1.2.1.3).

languages, for example, have rather complicated inflectional morphology and therefore are definitely not monosyllabic, and Japhug, a Gyalrongic language, lacks tonal contrasts (Jacques 2021).

Typological classification of Proto-Trans-Himalayan Given the striking internal typological diversity of Trans-Himalayan languages, the typological status of Proto-Trans-Himalayan is problematic. We have reasons to suspect that Proto-Trans-Himalayan is closer to Rgyalrongic in its typological features, actually. We note that the Rgyalrongic family and the Kiranti family are two distant groups within the Trans-Himalayan family (which is expected given that they are separated by the Himalayas) with seemingly cognate personal agreement affixes which is hard to be explained solely in terms of grammaticalization of pronouns, and one possible trace of personal agreement is also found in Tibetan (Jacques 2012; Jacques 2010; Jacques 2016). Considering the fact that the number of potential cognates between Rgyalrongic and Kiranti is around 150, and the number of cognates of them with Chinese is of the same order of magnitude,[3] it is highly likely that a Rgyalrongic-like personal indexation system originated at a quite early stage, which possibly is also the ancestor of Chinese (Jacques 2012; see Box 1.1 for more discussions). Another morphological device that probably has a Proto-Trans-Himalayan origin is the the **s-* prefix, which is also supported by reconstruction of Old Chinese. It sometimes is known as the causative prefix, although this usage is possibly due to grammaticalization of denominal verbs (Jacques 2015).

Box 1.1: Historical status of the personal agreement affixes

What exactly can be reconstructed to Proto-Sino-Tibetan is not without controversy. (a) We may assume that we have a Rgyalrongic-like polypersonal argument indexation system, which is a direct-inverse system, in Proto-Sino-Tibetan. (b) Or we can always argues that what can be reconstructed to Proto-Sino-Tibetan for certain is just personal affixes, as the underlying structure of a direct-inverse system is often not *that* different from more common alignments (Oxford 2017), and therefore the direct-inverse system itself possibly cannot be dated back to Proto-Sino-Tibetan and only the personal affixes can, which are the diachronic materials used to form a Rgyalrongic direct-inverse system in later stages. (c) As a step further, it can be argued that the personal agreement affixes were probably of pronominal origins, and hence we can argue for a morphologically impoverished Proto-Sino-Tibetan, and thus, the great typological distance between Proto-Sino-Tibetan and Chinese is no longer that striking.

The viewpoints listed above are critically examined in Jacques (2012). Although a comprehensive comparison is still lacking, Jacques notices that there exist agreement affixes in Rgyalrongic and Kiranti that appear to be cognates and yet cannot be easily explained by grammaticalization of pronouns and hence (c) cannot be completely right, thus at least implying certain verbal morphological complexity in a common ancestor. Further, by comparative analyses, he demonstrates that the verbal conjugation paradigms in Rgyalrongic and in Kiranti are remarkably similar (for example, only transitive verbs come with the aforementioned affixes), hinting to a biactantial agreement system in the common ancestor, and that Kiranti seems to retain certain residues of an inverse prefix, while the evolution of a direct-inverse system, despite theoretically possible, remains to be rare in Eurasia. The statement (b) above therefore is also not quite feasible, although not completely impossible. Anyway, a morphologically rich Proto-Trans-Himalayan is not controversial.

[3] See Shuya, Jacques, and Yunfan (2019) for an explicit historical comparison between Rgyalrongic and Sinitic.

1.2.1.3 Reconstruction of (pre-Classical) Old Chinese

Having established that Proto-Trans-Himalayan is morphologically rich, the evolution from it to today's largely analytic Sinitic languages is rather intriguing. There has to be a series of transitional stages between an ancestor language with somehow Gyalrong-like typological traits and the modern Sinitic family.

Based on philological evidence, the early history of Chinese is commonly divided into an Old Chinese period containing both the pre-Classical period (§ 1.2.2) and a Classical period from the Spring and Autumn period (§ 1.2.3) to late Han dynasty (§ 1.2.4), and a Middle Chinese period. The way Middle Chinese works is already largely comparable to contemporary Sinitic languages. Reconstruction of the Chinese language during the Old Chinese period therefore is essentially for understanding the emergence of modern Sinitic features.

Materials and challenges The Old Chinese period is quite long, spanning more than 1,000 years. Conventionally, the term *Old Chinese reconstruction* is mainly about the reconstruction of the language in the *pre*-Classical period, and therefore a full survey on the progress of Old Chinese reconstruction beyond the scope of this work. Understanding how pre-Classical Chinese probably looked like *as a language* and not just as dead texts however may provide some insights for understanding the origin of certain features of Classical Chinese, like how valency alternation works.

The most conventional pre-Classical materials with phonological information include information contained in the phonetic components of Chinese characters, both the transmitted ones and ones in excavated texts, and the rhyme patterns in the *Book of Odes*. Less conventional but no less important materials include evidence from modern Sinitic languages, especially Min, which contains features that deviate from Middle Chinese rhyme books, and evidence from early Chinese loanwords to other languages. Any reconstruction must be compared with Middle Chinese evidence and show allow natural explanations of the Middle Chinese phonology (Baxter 2014, Ch. 2).

None of the materials is alphabetic, and Old Chinese reconstruction is expectedly a chaotic field with lots of controversies. Baxter (2014) represents an important milestone in this field, making use of all the materials mentioned above as well as a rich toolbox of philology, the historical comparative method, and internal reconstruction. But even it has severe flaws (Harbsmeier 2016). To be fair, currently we do not even have a well-defined target of reconstruction. Pre-Classical texts show wide variation in syntax and presumably in phonology, and the Chinese scholar tradition clearly states that *Book of Odes* consists of poems collected at different times and different locations (§ 1.2.2; Harbsmeier 2016, pp. 478-480). In the Classical period, texts made it clear that mutually incomprehensible Chinese varieties existed (Harbsmeier 2016, pp. 446-447). A distinction still needs to be made between Old Chinese(s) and Proto-Sinitic, contrary to the claim of Baxter (2014, p. 2). Finally, even when we are talking about a relatively well-defined branch of Old Chinese, for instance the syntactically largely unitary Classical Chinese discussed in this work (which could be distant from the actual spoken Sinitic dialects; § 1.2.3), there is no guarantee that it has a unitary phonology (Harbsmeier 2016, pp. 488-490).[4]

[4]It should be noted that the problem is not restricted to Old Chinese. Strictly speaking, that there was a unitary Middle Chinese as reflected by rhyme books and rhyme tables also should not be taken for granted. What rhyme books and rhyme tables were intended for could probably never known for certain: they could be prescriptive works documenting the mixture of several respected dialects, or even an early Neogrammarian-ish reconstruction attempt. Besides, we are also not really sure about

Due to all of these issues, we will avoid presenting a fine-grained reconstruction of Old Chinese (Classical or pre-Classical) phonology in this work. Below, we list some widely accepted features of (pre-Classical) Old Chinese to give the reader an intuitive and vague feeling of how it feels like.

Origin of tones In Middle Chinese rhyme books, there are four tones reported: the LEVEL tone (平聲), the RISING tone (上聲),[5] the DEPARTING tone (去聲), and the ENTERING tone (入聲).

Let us first focus on the ENTERING tone. It can be easily seen that rhymes falling under the category of the ENTERING tone are much fewer than rhymes pertaining to the other three tones, with a lot of nasal rhymes missing. The reflex of roots with ENTERING tone in modern Sinitic languages, if any, all end with a stop (consider e.g. Cantonese *-p*, *-t*, *-k*). These words, when borrowed into Japanese, usually became disyllabic. These facts can be explained perfectly by assuming that in Middle Chinese, they ended in *-p*, *-t*, or *-k*, and this seems to be the defining feature of the ENTERING tone, which, in Middle Chinese, was strictly speaking not a tone at all.[6]

Now given that modern Sinitic tones can be easily explained by conditional sound changes starting from three tones (Sagart 1999a), the rest three tones in Middle Chinese were real tones. But we can try to apply the methodology inspired by the analysis above to Old Chinese, and the conclusion is that there was probably no tone in the Old Chinese period. Below, I briefly repeat the argumentation in Sagart (1999a).

The origin of Vietnamese tones has been demonstrated to be final consonants. That in the oldest layer of Chinese loan words into Vietnamese, words in the Middle Chinese DEPARTING tone regularly correspond to Vietnamese words in a tone originating from the final consonant *-s* strongly suggests that in Old Chinese, the origin of the DEPARTING tone is also something like **-s*. This is supported by textual evidence: in early Chinese transcriptions of foreign words, the DEPARTING tone sometimes corresponds to foreign *s*, and some early Chinese load words into Korean even retain the *s*. Therefore, the DEPARTING tone is likely a reflex of the Old Chinese word ending consonant *-s*.

In a similar way, the RISING tone is related to an earlier glottal stop based on evidence from Sino-Vietnamese words. We also note that RISING tone words were preferred to transcribe Sanskrit short vowels, which is consistent with the glottal stop assumption, as the latter makes the syllable sound short. Thus, the origin of the RISING tone can be tentatively reconstructed as **-ʔ*.

Now that all other tones have been reduced to syllable final consonants, the remaining LEVEL tone can be simply understood as the "default" case, i.e. open syllables or syllables ending with nasals. We have thus shown that Old Chinese lacks tones.

---

the phylogenetics of modern dialects: if we are to stipulate a Middle Chinese *node* in the Sinitic family tree which is the common ancestor of all modern Sinitic dialects, then how did it emerge from the messy congeries of dialects during the Middle Chinese *period* (Harbsmeier 2016, pp. 477-478)? We are more confident with the notion of a unitary Middle Chinese (defined in the rhyme books) as the common ancestor of modern topolects, mainly because comparative works do suggest that the former is indeed a good starting point to analyze the phonology of the latter. Still, Neogrammarian comparative studies based on modern topolects are urgently needed, not mentioning the fact that many topolects are quickly going extinct without proper documentation.

[5] The term may be confusing for a Modern Standard Mandarin speaker, as its reflex in Modern Standard Mandarin is the third tone, while the "raising" tone in Modern Standard Mandarin is the 陽平 tone, which historically was from the LEVEL tone.

[6] In modern Cantonese there are actually three reflexes of the ENTERING tone and they are real tones, but their emergence can be explained by

Affixation The fact that Chinese tones probably originated from word-final consonants and the fact that in reading traditions of Classical texts, tone change represents subtle modification of the meaning of a root (known as 四聲別义 'four tones distinguishing between meanings') immediately suggest that Old Chinese has derivational suffixes. All derivations that change the tone of a word to the DEPARTING sound, for instance, can now be explained by a **-s* suffix. Several prefixes have proposed as well, based on alternations of the onset, e.g. multiple **s-* prefixes (Baxter 2014, pp. 53-59).

Initial consonant clusters The syllable structure of Modern Standard Mandarin and is relatively simple. The syllable structure of Middle Chinese was slightly more complicated, allowing more syllable-final consonants (i.e. ENTERING tone discussed above). However, all existing materials point to a much more complicated Old Chinese syllable structure.

A popular perspective, although not universally accepted yet, is that consonant clusters were prevalent in Old Chinese. This can be supported by 諧聲 *xiéshēng* relations, i.e. semi-homophonous relations observed in Chinese characters without semantic relations that share a component. Based on inspecting these phonetic series of characters in Modern Sinitic languages, we conclude that what is shared is likely the *initials* of the syllables. Now, consider the semi-homophonous series 监, 槛 (pronounced as *jiàn* in the reading convention of Classical texts). Then consider another series 滥, 蓝. The fact that they share the same phonetic component 监 strongly suggests that the four characters belong to the *same* series when the characters were first invented. A reconstructed initial like **kl-* therefore may be proposed for the phonetic value of the phonetic component 监. A similar "broken" series include 禀, 凛 (something like **pl-*). And now we find a class of consonant cluster initials roughly with the structure of **Cl-*, which actually is reconstructed by Baxter (2014, pp. 50-51) as **Cr-*.

Final consonant clusters Consonant clusters at syllable finals are also proposed. First, Qing dynasty scholar 段玉裁 Duan Yucai noticed that in *Odes*, rhyme sequences in the LEVEL, RISING and ENTERING tones can be isolated (meaning that all rhymes in a poem have the same tone in Middle Chinese), while the rhyming of DEPARTING tone words appears to be random. His conclusion was that there was no DEPARTING tone in *Odes*. Now we know that there was probably no tone at all at least in pre-Classical Old Chinese, and Duan's observation means that the final consonant **-s* corresponding to the DEPARTING tone was likely almost always not a part of the root, but something like a suffix. But if a word with the ENTERING tone (which means it ends with **-p*, **-t*, or **-k*) rhymed with a word that ends with **-s*, then the latter should also contain **-p*, **-t*, or **-k* near its end: and hence the **-ps*, **-ks*, **-ts* final consonant clusters (Hill 2019, p. 183).

In Middle Chinese, there are four rhymes (祭泰夬废) in the DEPARTING tone that do not have counterparts in the LEVEL and RISING tones, and words with these rhymes having semi-homophonous relations with ENTERING words, based on *xiéshēng* relations seen in their characters. This means these rhymes probably ended with **-ps*, **-ts*, **-ks* in Old Chinese, and have both ENTERING and DEPARTING features. Consonant clusters appear here again.

We also note that there are RISING tone words with nasal finals in Middle Chinese: 上 itself is an instance. This suggests a consonant cluster roughly with the form of **-ŋʔ* at the end of the words, which probably should be **-ʔŋ* to make it looks typologically feasible. We note that the form **-ʔŋ* can even explain the origin of the RISING tone, as it makes the pronunciation *weak* in the middle, which naturally is close to a

pronunciation that is *low* in the middle, and hence the modern RISING tone.

Minor syllables The format of the *Odes* strongly suggests that the language(s) behind the poems were monosyllabic. Still, we have evidence for residues of minor syllables before the main syllables, known as *loose* pre-initials in Baxter (2014); their *tight* counterparts are ordinary consonant clusters. In *Odes*, some rather strange forms, like 帝命不時, contain 無 that would otherwise be interpreted as a negator, but the interpretations of them contain no negative meanings according to commentaries. Not all of them can be naturally explained as rhetorical questions. We can also find records like 聿……吳謂之不律 in 说文, further suggesting the existence of loose pre-initials (Hill 2019, pp. 122-1235). The reconstruction of these pre-initials is still controversial, but their existence at least in some historical stages and areas seems to be beyond reasonable doubts. The syllable structure of pre-Classical Old Chinese therefore should be (C*ə*)CVC.

Word order Texts during the Old Chinese period do *not* have a rather flexible word order, as is shown in the rest of this work. The evolution towards a largely analytic syntax therefore had to already have started at the Old Chinese stage.

1.2.1.4 How Chinese acquired its typological features?

To summarize what has been revealed in reconstruction works, absence of tones, words with minor syllables, and various affixation mechanisms were characteristics of Chinese at a certain stage (Sagart 1999b, p. 13). In the earliest Chinese texts, however, we already see indications of emergence of stereotypical Sinitic features: the syntax was already quite rigid, the number of available affixes was limited and their productivity was likely weakened, and the syllable structure was under simplification (TODO: ref). The factors pushing Chinese on its path to its appearance today are of interest.

Relation with Chinese characters It is sometimes argued that the loss of morphology is due to the Chinese characters, which do not record inflection well. This claim is linguistically untenable, first because most people before modernity were illiterate, and second because it is quite straightforward to dedicate certain characters to represent inflection (which is indeed how the *kana* system used to write Japanese developed). A reasonable, but not consensually accepted hypothesis is that it was due to language contact, which is discussed in § 1.2.1.5.

Area typology Certain languages in the Trans-Himalayan area that do not have clear genetic relations with the Trans-Himalayan family, on the other hand, do have the purported Trans-Himalayan typological features. Kra-Dai languages have these typological features as well (Sidwell and Jenny 2021, p. 434), and yet the most likely relatives of them are the Austronesian family, which do *not* have these typological features at all (Sagart 2004; Ostapirat 2005).

The evolution of tonal distinctions happens in Kra-Dai as well (Sagart 2019).

1.2.1.5 Language contact

The makeup of the Old Chinese lexicon Less than a half of Old Chinese words have clear etymological Tibeto-Burman counterparts, and the rest of the lexicon comes from Austro-Asiatic and Tai contributions (McCraw 2010). Although in later periods, Chinese (at different stages of development) was the *origin* of borrowed terms, in the

prehistoric era, Chinese was frequently the *recipient*, not the *donor*.

Neighboring languages Certain records in the Classical period indicate the existence of non-Sinitic populations in southern kingdoms. For example, the 越人歌 *Yuèréngē* '*Song of Boatmen of Yue*', a song of boatmen preserved in *Garden of Stories* that was both transcribed in Chinese characters (in a sort of Old Chinese pronunciation) and translated, was proposed to be composed in a Kra-Dai language (Zhengzhang 1991). What we are discussing above happened long *before* the Classical period, but the relevant populations did not come out of no where and should be there as well when Proto-Sinitic was being formed. Further evidence supporting this claim includes

The mechanism for the mixing between a more Trans-Himalayan version of Proto-Sinitic and various Southeast languages may be the follows. The Shang dynasty, which left us with Sinitic documents (§ 1.2.2.1), did not come out of nowhere, and early state formation, dominated by speakers of some sort of Proto-Sinitic. The It is possible that Proto-Sinitic was spoken as a lingua franca, used by

Syntax of Old Chinese and language contact A natural idea is that the syntax of Old Chinese – and hence the syntax of Classical Chinese – was also influenced by non-Trans-Himalayan languages. The SVO word order, the prepositions, and clause-initial subordinators all stand in contrast to Tibeto-Burman languages. On the other hand, Old Chinese has sentence-final particles, which seem to be more consistent with a SOV order, and occasionally we observe relics of a SOV order as in 不我知也 (Ch. 2, 16); these features are not surprising within the Trans-Himalayan family.

How sentence-final particles were grammaticalized is not always clear: in oracle bone inscriptions there are already some SFPs; see http://ir.lib.scu.edu.tw:8080/bitstream/987654321/11005001.pdf: 句尾亦偶有少數的語氣詞

We note that the all of the Southeast Asian groups with SVO syntax seem to have this order as far back as we can trace, and therefore in the aspect of word order (and probably other aspects), Chinese is the *recipient*, not the *donor* (DeLancey 2013, p. 82-85).

1.2.2 Pre-classical period

1.2.2.1 Oracle bones

The earliest attested Sinitic texts were oracle bone inscriptions from the Shang dynasty. It should be noted that people during the Shang dynasty *wrote* (and not just *inscribe*) as well: the counterpart of character 册 exists in oracle bones as well, which looks a bundle of bamboo and wooden strips. Still, the only materials that reflects the language of Shang currently available are inscriptions.

The characters in these inscriptions are structured in a way largely comparable to later Chinese characters. Since the phonetic values of characters in oracle bone inscriptions are not self-evident, some may even question whether these inscriptions are Sinitic at all: some may argue that the language of Shang was a (possibly unclassified) SVO language, and when the Zhou dynasty was established on the ruins of Shang, its culture was borrowed, including both the writing system and the SVO syntax. This hypothesis is not quite plausible, because in such a scenario, usually what happens is *lexical* borrowing, and the introduction of a writing system should be minimal to the structure of the language (DeLancey 2013, p. 88). The introduction of Chinese characters to write Japanese is a good reference point. Further, phonetic borrowing (假

借, 'borrowing, i.e. a character originally representing one morpheme to represent another morpheme with a close sound') in oracle bone inscriptions neatly fits into reconstructed Old Chinese phonology (e.g. Baxter 2014, pp. 115).

The language of Shang in certain aspects is closer to the

1.2.2.2 Post-Shang texts

Oracle bone inscriptions are a 20th century archeological re-discovery not known to Classical Chinese authors. Zhou's replacement of Shang seemed to be accompanied by a great cultural revolution: the last king of Shang, namely King Zhou of Shang, was known by his arrogance, his corruption, and his tyrannic reign. Yet no one seems to remember the arguably most scary aspect of Shang dynasty: human sacrifice. To Classical authors, the earliest available texts are documents preserved in 《尚书》(lit. 'venerated documents'), often known as the *Book of Documents* in English. Since these texts are from ancient kings whose deeds were romanticized by Confucian scholars, these texts were highly venerated and yet deemed as 诘屈聱牙 ('twisted, hard to pronounce') by post-Classical authors.[7] They were something that had to be read with commentaries, the latter written in easier Classical Chinese. These documents therefore should be regarded as pre-Classical, although they did contribute sporadic phrases and grammatical words (e.g. the copula 惟 or the pronoun 厥) that were occasionally used in Classical Chinese works as a way to polish an article.

One thing worth mentioning, which is also a part of the cultural paradigm shift after the end of the Shang dynasty, is that the language of the *Book of Documents* and the language of oracle bone inscriptions are clearly different. The most notable fact on this aspect is that the aforementioned pronoun 厥 appears frequently in the *Documents*, but it appears neither in oracle bone inscriptions nor in Spring and Autumn works. Presumably, *Book of Documents* contains predominantly early Zhou dynasty texts, while oracle bones dates back to Shang, and the differences we are observing reflect dialectal differences between the ruling classes of the two dynasties.

Another fairly early source is 《诗经》(lit. 'poem classics'), also known as the *Book of Odes*, which contains poems dates back to as early as early Zhou. We note that the *Odes* is often considered Classical, especially in a Confucianism context, and yet given its poetic nature, its influences to Classical *proses* are not direct – and the grammar of poetry varies wildly in different periods. This work mostly concentrate on the grammar of proses.

The languages of these old documents were more or less different from the prototypical Classical language, which is introduced in the next section.

1.2.3 Spring and Autumn and Warring States

The majority of texts that shaped Classical Chinese proses were written in a time when Zhou was already substantially weakened. This period that witnessed prolificacy of Old Chinese works can be divided into two periods: the Spring and Autumn period which was filled with chaotic (but not intense) wars between numerous dukedoms, and the Warring States period which observed intense wars between seven major states, resulting in a unified Qin empire, which soon broke down because of resistances to its barbaric policies and eventually was superseded by Han dynasty (§ 1.2.4). The

[7] For example by Han Yu in 《进学解》(*Analysis of academic advancement*).

language of this period diverges tremendously from the pre-Classical period. For example, the copula 惟 had died out of use and the copula construction had been largely replaced by the nominal predication construction (§ 2.2.1). The conjunction 而 is never seen in pre-Sprint and Autumn texts but had already made its way into the *Analects*. The lexicon also underwent huge changes.

> **Box 1.2: Lexicon change**
>
> List some lexicon changes.

There are clues suggesting regional variances. Students of Confucius noticed that when he recited Classical texts and presided rituals, he used 雅言 or 'elegant speech' (*Analects* 7:18). This suggests a possible diglossia at as early as Confucius's own age, with the "elegant speech" conceivably being the language of intellectuals of Zhou Dynasty. Comparison between the language of Classical proses and the language(s) of poetry shows the relative homogeneity of the former, while the latter both demonstrate divergence from the language of the proses and regional differences.

> **Box 1.3: Peotry and prose**
>
> This is presumably due to how the texts were transmitted. It is likely that they were passed by recitation, and regularization happened to proses when there was a predominant dialect, while the prosody and rhyme structures of poems efficiently locked them to their original forms.

The language of 楚辞 (*Verses of Chu*), for example, has the following differences with the language of the proses. The first is a Kra–Dai substrate.

> **Box 1.4: Chu dialect**
>
> Find references.

The language of the *Odes* also seems to slightly deviates from
Dialectal differences have also been observed within the *Odes* (List et al. 2017).

1.2.4 Han dynasty

The last batch of uncontroversially classical works were composed during Han dynasty, among them the most important being *Records of the Grand Historian*. The language of *Records of the Grand Historian* shows notable but largely qualitative differences compared with earlier historical works, the most important one being 《左传》. Notable changes include more pre-verbal adverbials, reduction of prepositional verbs, regularization of constituent orders, and also proliferation of disyllable words It is therefore suggested that Han dynasty texts and pre-Qin texts reflect two stages of post-Zhou developments of Chinese, although the change was definitely not as radical as the change from the *Documents* to Spring and Autumn texts (He 2005, pp. 260-264).

1.2.5 Post-Classical periods

The end of Old Chinese – and hence the end of the classical period – is marked by the collapse of the case inflection in the personal pronoun system, the emergence of 是 as a copula (and not just a demonstrative), the appearance of the disposal construction (i.e. the 把 construction) and the so-called long passive construction.

Box 1.5: References for Middle Chinese and modern Mandarin

- James Huang
- etc.

Expectedly, despite purification attempts, vernacular elements made their ways into not only administrative documents but also pure literature and scholar works. Classical Chinese or 文言, in the broadest sense, is a term that covers all genres whose grammars are roughly based on the Old Chinese canon but may have a varieties of innovations.

Box 1.6: Late regularization attempts

韩愈、唐宋八大家、因明学

1.3 About this work

1.3.1 Theoretical framework

The theoretical framework of this work is essentially Distributed Morphology (for an introduction see Siddiqi (2009)) plus Cartographic Syntax (e.g. Cinque (1999)), although I intentionally choose to reuse the terminology in descriptive grammars (see below).

The architecture of grammar is assumed to be in line with the basic assumptions of Distributed Morphology (§ 2.1), where we have a list of roots (List A), each of which is only compatible with certain syntactic positions in post-syntactic phonological realization (List B), and grammatical objects – bundle of roots and functional heads, or even larger objects – can be lexicalized with custom meanings (List C). Lexicalization or in other words *semantic* fossilization is important for certain aspects of Classical Chinese grammar (e.g. § 7.2.2), which however can be well captured within Distributed Morphology, without lexicalist assumptions (cf. Bruening 2018). We note that lexicalization of a complex structure may eventually lead to the collapse of its internal makeup, causing *syntactic* fossilization. An example is the collapse of verbs taking complement clauses into compound verbs. The end point of fossilization is a synchronic *root*, on which only diachronic analysis is possible. Syntactic fossilization is important in Classical Chinese due to its long history.

The analyses in § 2.2 and § 2.3 are inspired by the extended CP and DP structures proposed in Cartography. To avoid confusion caused by technical terms in generative syntax, I intentionally use terms like *sentence*, *nucleus clause* and *argument structure* in place of CP, TP and vP. Further, the notion of functional heads should be avoided, and concepts like SpecTP and SpecvP have to be replaced by concepts like *clausal subject/pivot* and *subject in the argument structure* (§ 2.2.5), and we should talk about *noun phrases* instead of DPs. Similarly we cannot talk about DO or CAUSE light verbs; I replace these concepts by concepts like *DO clause* or *CAUSE clause* (§ 2.2.3). The "core" of CPs and DPs (i.e. the roots at their centers plus the categorizers) – instead of the grammatical markers – should be known as heads. After doing so, we rediscover the good old subject-predicate and verb-object relations, subordination and coordination structures, and other *constructions*, as well as *grammatical relations* within them.[8] This procedure has been demonstrated in Deng (2010), which shows that Minimalist generative syntax and the constituency-based analysis strongly inspired by American structuralism

[8]Here the term is used *without* the implication that a construction is somehow understood as a whole and its internal structures should not be further analyzed, contrary to the fundamental hypothesis of e.g. various Construction Grammars.

and is exemplified in Huddleston and Pullum (2002) and works outlined in § 1.4.3 are compatible to each other, with the former being a more concise form of the latter and the latter being a logical consequence of the former under certain assumptions.[9]

Grammars can be written in terms of *dependency relations*, instead of *constituency relations*. The two however are largely equivalent (Boston, Hale, and Kuhlmann 2009). In his Basic Linguistic Theory, Dixon (2009) fervently argues against constituency analysis (and also other aspects of generative syntax) and advocates for a "flat" constituency structure, possibly with the levels of clause and noun phrase only, where the rest of the grammatical information is represented by dependency relations (e.g. Dixon 2009, p. 49). Yet the primacy of noun phrases and clauses in syntax is indeed acknowledged by generative syntax (for instance they are *phases*), and the binary constituency relations in generative syntax and in American structuralism have consequences. For instance, when applied to the argument structure, they are related to extractional properties of arguments in valency alternation (§ 2.2.5.1), and the subject-predicate binary division is directly related to the clausal pivotal status of the subject. These phenomena of course have to be taken into account by Basic Linguistic Theory, and labels like "clausal pivot" and "surface S, deep O" have to be attached to dependency arcs between the verb and the arguments, essentially labeling the "distance" between the two. As a parallel, the ancient India grammarian Pāṇini initially proposes a grammatical framework in which arguments are all equal, but later commentators still effectively set up a pivot position in the argument structure (Box 1.7). Other phenomena that can be directly include ordering of tense and aspect auxiliary and adverbs à la Cartography, which is attested in languages completely unrelated to Germanic or Romance languages that have been thoroughly investigated in Cartography (e.g. Grimm 2021, pp. 166-167), and also available clause sizes: it is rare that in a language with tense, aspect and modality marking, there exists a clause type with discourse devices but no tense, aspect and modality marking, which can be explained simply by the fact that CP is built on TP.

Besides the theoretical problems outlined above, there are also some minor, largely notational inconsistencies between the grammatical theories mentioned here. One such inconsistency is the definition of the *phrase*. Dixon (2009) calls the main verb plus auxiliaries – without any argument – as the *verb phrase*. This actually makes sense in generative syntax because arguments are phases themselves, and the tense, aspect and modality categories marked by the auxiliaries are in some senses closer to the main verb. Yet the term *verb phrase* generally means the verb plus internal arguments in constituency-based analyses. Another problem is the definition of *word*. Since we reject the lexicalist hypothesis, we need to distinguish between phonological wordhood, morphological wordhood (the boundary of the morphological template is the boundary of the morphological word), and syntactic wordhood. Syntactic wordhood in turn has several definitions. We can define a word to be a very small constituent: if it is impossible to infer the argument structure of a compound verb in a given language,

[9]Note that Deng (2010), Huddleston and Pullum (2002) and works in § 1.4.3 are all lexicalist, which we have argued is not necessary to account for phenomena purportedly supporting the lexicalist hypothesis. These works are also based on the more traditional X-bar theory: the syntactic differences between the subject and the object, for example, are explained by the difference between specifier and complement. In Cartography, what is recognized as a complement may be a lower specifier: thus [V O]$_{V'}$ is replaced by a "transitivity phrase" [O [TRANS √V]$_{TRANS'}$]$_{TRANSP}$ (Siddiqi 2009), where the categoryless verb root is first combined with a transitivity functional head and then combined with the object. The resulting tree diagram is quite like the X-bar scheme before, say, the DP hypothesis: the only difference being that there are many X'-like projections, all of which may appear in other constructions. For instance, the English *nominal* mentioned in Huddleston and Pullum (2002), which is often recognized as N', can actually be a modifier in the DP, as in *a [Fuji apple] tree*. Deng (2010) shows that the subject-predicate structure or the verb-object structure can all be derived from the traditional X-bar scheme: the Cartographic approach pushes this decomposition even further and demonstrates that the traditional X-bar scheme itself can be further decomposed into a series of functional heads and roots.

then we conclude that the two branches of the compound are not categorized, and therefore the compound is a rather small constituent and hence a word. But in this way *sinned* is *not* a syntactic word as it involves a clausal category (i.e. the past tense). For *sinned* to be a syntactic word, syntactic wordhood can be based on Dixon's verb phrase, and inevitably *have been eating* is a syntactic word.

In conclusion, we maintain that Chomskyan generative syntax, American structuralism-styled constituency analysis as in Huddleston and Pullum (2002), and Basic Linguistic Theory (the de facto unified framework in modern linguistic description of underdocumented languages) are coherent and can be seen as three "representations" of the same grammatical complexity class, and their differences are mostly notational.[10] Which framework to use is to be determined by the properties of the language. For instance, although the definition of the verb phrase does not alternate the grammar system substantially, Dixon's definition works more smoothly for a language with a lot of auxiliaries but rather infrequent subject-sharing coordinations. It turns out that for most constructions, the method in Huddleston and Pullum (2002) is a good choice for Classical Chinese.

1.3.2 Coverage

Contemporary grammatical description typically starts in a topic-by-topic way: the first grammar of a language likely outlines how a head noun is modified by an adjective or another noun, while the relative order of different modifiers in the noun phrase and their scopes are typically skipped. To some degree, such a strategy may be described as an *ethno-philological* approach: it prepares the reader to comprehend *natural texts* in the language being described, as complex structures, after all, tend to appear less frequently than simpler ones. Classical Chinese is a classical language, and expectedly, philology-oriented works have dominated the field of grammatical research on Classical Chinese.

This work, on the other hand, is an attempt to study the grammar of Classical Chinese as a machine that takes lexical items and produces *arbitrarily* complex utterances, and to understand the structure of this machine. The approach is admittedly inherently problematic for a dead language, which has no native speaker with acceptability intuitions. Further, as is outlined in § 1.2, texts that are considered to be define Classical Chinese have great internal diversity. The grammatical system presented in this work therefore resembles what Huddleston and Pullum (2002) calls *International Standard English*, i.e. the shared core of all important contemporary varieties of English.

It should be noted that the existence of such a shared Classical grammatical core should not be taken for granted in all Classical-looking texts. During the Republican

[10] A more important controversy is the mental status of grammar. The position of this work treats grammar as a semi-autonomous component of human's cognitive abilities. But it has been argued that grammatical constructions originate from domain-general cognitive abilities, and there is actually no such thing as an autonomous mental grammar. Detailed discussions on this topic are far beyond the scope of this work, and can only be finally settled down with the assistance of neurological studies. Here we just note that currently no comprehensive description of a language has been successfully attempted under this line of thinking. On the contrary, in physicists' terms, the grammatical framework adopted here is at least a good *effective theory*.

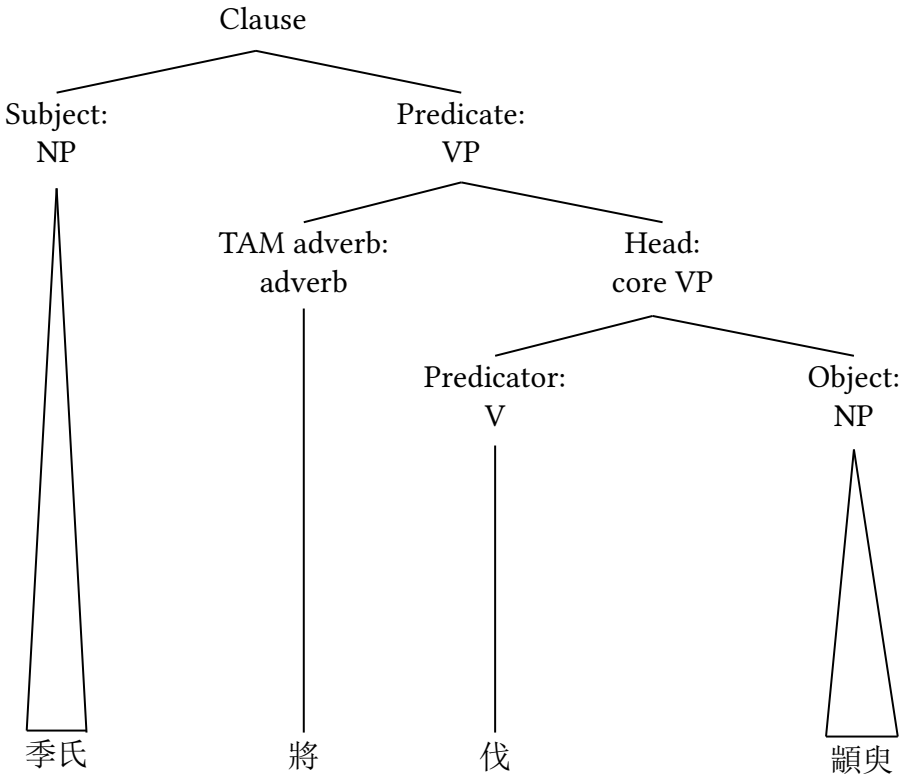
Another topic is whether natural language-like "grammars" appear in systems that are demonstrably independent to the human neurological linguistic capacity, which may also challenge the domain-specific status of the latter. This is partly discussed in § 3.1. Note however that such discussions are largely irrelevant to linguistic description, for reasons mentioned above.

period of modern Chinese history, for example, some official documents were written in a pseudo-Classical style: the abstract syntax behind these documents can well be captured within the framework of modern Mandarin, but the possessive marker 的 is replaced by 之, and the relative marker 的 in headless relative clauses is replaced by 者 (不服管教的 'who does not conform to instructions' is replaced by 不服管教者). What is Classical in these documents is primarily the superficial forms. Still, because of the highly formulaic nature of these documents (for instance, usually no sentence final particle appears, and valency alternation besides the passive is discouraged), it is also likely that these documents can be considered as Classical in syntax as well. More details can be found in § 11.2.2.

1.3.3 Notations

In § 1.3.1, we have mentioned that the framework of Huddleston and Pullum (2002) seems to be a good starting point of the description of Classical Chinese. An example of the tree diagram representation of the constituency analysis (e.g. Huddleston and Pullum 2002, p. 954 [9]) is shown in Fig. 1.1. This representation does have one problem: it does not clearly distinguish function words from content words, while the distinction has syntactic consequences (see e.g. § 2.2.4 and § 2.2.3.2.1).

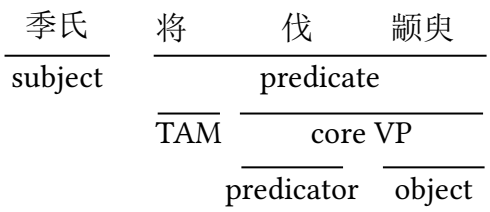
Figure 1.1: Tree diagrammatic analysis of 季氏將伐顓臾 'Jishi (a nobel family which controls the politics of the Kingdom of Lu) is going to invade Zhuanyu (a vassal state of Lu).' (*Analects*)



Researchers in modern China conventionally draw lines below constituents with labels to represent constituency relations. Fig. 1.2 contains exactly the same information in Fig. 1.1 and is more compact. It also highlights the lexical/functional distinction.

For instance, if 將 were a future tense particle and not an adverb, we could refrain from giving a label to it (and eliminating the "TAM" layer in Fig. 1.1), implying that it is a grammatical marker and not a (content) constituent of the predicate.

Figure 1.2: More compact version of Fig. 1.1



The problem of Fig. 1.2 is also its compactness: the label of a constituent cannot be too long, or otherwise the whole diagram cannot be fit within one line. In this work, we represent examples in a way that is more aligned to modern description grammars: we place constituents in brackets, and then label these brackets by subscripts. (1) is an example. When we have grammatical markers, we do not put them (e.g. the sentence final particle 矣 in 2) in brackets to highlight the lexical/functional distinction, consistent with the suggestion of Dixon (2009, p. 49).

(1) [[季-氏]$_{\text{subject: NP}}$ [[將]$_{\text{TAM adverb}}$ [伐 顓臾]$_{\text{core VP}}$]$_{\text{predicate: VP}}$]$_{\text{verbal clause}}$
Ji-family.branch.lineage.name will subdue Zhuanyu
'Jishi is going to invade Zhuanyu.' (*Analects*)

(2) [[[民$_{\text{subject: NP}}$ [鮮 —$_{\text{object}}$]$_{\text{predicate: VP}}$]$_{\text{subject: complement clause}}$ [久]$_{\text{predicate: AdjP}}$]$_{\text{verbal nucleus clause}}$ 矣$_{\text{SFP}}$]$_{\text{sentence}}$
people lack long SFP
'It has been quite long that people lack this!' (*Analects*)

Note that in (1, 2) we do not transcribe texts written in Chinese characters into any alphabetic form. This is a practice *against* contemporary common practices in descriptive linguistics. Just like Latin, historically, Classical Chinese was read by different readers with different phonological systems. The texts that set standards for Classical Chinese were excerpts from several closely related varieties of Old Chinese, which were likely spoken languages when the texts were written (§ 1.2), but reconstruction of the phonology of Old Chinese remains a highly debated topic, and no mutually accepted lossless romanization that reflects the historical phonology exists currently. Alternatively we may choose to use the pronunciation of modern Standard Mandarin, but this does not give us much insights into the phonological reality of these texts. Therefore, when doing interlinear glossing, we keep the Chinese characters. When translating texts into English, modern Standard Mandarin *pinyin* (i.e. the standard romanization) is used for proper names. This is however just for convenience, as the *pinyin* representations of these names also appear in the gloss.

1.3.4 Organization of the book

We adopt a top-down order in writing this book. The starting point is the "philological", surface-oriented enumeration of the linear templates of attested clauses and noun phrases in § 2.2.1, § 2.2.2, and § 2.2.2. Grammatical systems identified when analyzing and comparing surface forms are then introduced in the rest of the sections of § 2.2 and § 2.3, based on which a lexical parts of speech distinction is made. If a section in Ch. 2 is too long, it is moved to chapters following Ch. 2.

This organization is in contrast with many descriptive grammars, which start with e.g. the nominal and verbal morphology, and then discuss grammatical relations like possession, modification, and argument structure, and then outline the structure of the simple clause and complex clauses. This is not the approach taken here, partly because of what is discussed in § 1.3.2, partly because Classical Chinese is rather analytic in its grammar, and a well organized study of its grammar has to have the whole syntactic environment in mind: otherwise it will likely collapse into a dictionary of the various usages of function words. This is exactly what happened in the native grammatical tradition of Classical Chinese (§ 1.4.1).

1.4 Previous studies

1.4.1 Native grammatical traditions

Classical Chinese authors had conversations about grammaticality and uses of grammatical particles reminiscent of how English native speakers with some exposure to the study of English grammar but no formal training: "delete the *the* here and your sentence looks more concise". No attempts were made to establish intermediate concepts between words and utterances, like structural templates of phrases or grammatical relations, and to organize the grammar as a machine producing acceptable utterances: discussions on grammatical topics were either for education or for rhetorics.

The grammatical awareness of Classical Chinese authors was somehow comparable to what an ancient Roman grammarian or *grammaticus* did, who set his main role as a secondary educator, refrained from analyzing some sort of "underlying" or "internalized" system behind the surface forms and was satisfied by mostly surface-oriented patterns, and would not set up any intermediate concepts between the word and the utterance (Matthews 2019, pp. 7,35,47-48). On the other hand, this approach is contrary to the practice of the Paninian Sanskrit grammatical tradition, which, in today's terminology, starts with dependency relations and abstract features (§ 1.3.1; Box 1.7) and uses a set of morphophonological rewriting rules to produce the corresponding surface forms (Kiparsky 2009).

> **Box 1.7: Where does Pāṇini disagree with modern linguists?**
>
> The main difference between Pāṇini's treatment of Sanskrit and modern linguistic theories is that Pāṇini apparently treats all dependency relations equally and there is, for example, no concept of the pivot or the "external argument" of a clause. This is however modified in the commentaries of his *Aṣṭādhyāyī*, which explicitly allows an argument being promoted to the agent position because of the intentions of the speaker (Keidan 2017). The agent position thus becomes a subject position at least in the argument structure, consistent with modern practices (§ 2.2.5.1).

The Paninian tradition therefore is extremely close to modern linguistic description practice; the most important difference probably is that modern linguistic description, practically, may even be less rigorous than *Aṣṭādhyāyī*, because of possible competing "mind grammars" among speakers with mutual intelligibility or even within the mind of one speaker, and also the fact that a description as detailed as *Aṣṭādhyāyī* requires corpus data whose quality and quantity exceed the capacity of most field linguists.

The Classical Chinese grammatical tradition appears even looser compared with the Roman tradition in that the former did not even attempt to recognize parts of speech; this however was deeply rooted in the structure of Classical Chinese in that

Box 1.8: Ancient Chinese grammatical tradition and Roman tradition

Is the situation somehow close to what a Roman grammarian (*grammaticus*) would do? It seems that Roman grammarians also didn't care about abstract structures. See:

- Use and Function of Grammatical Examples in Roman Grammarians
- Quintilian's 'Grammar' (Inst.1.4-8) and its Importance for the History of Roman Grammar
- What Graeco-Roman Grammar was about

On the other hand, phonology was an active topic in ancient China. This was probably due to the influence of

1.4.2 Perspectives of European missionaries

Large-scale, systematic and reliable grammatical description of Classical Chinese had unfortunately been largely lacking for quite a while, and until recently, when Classical Chinese was studied linguistically, it was the historical phonology that was studied (Pulleyblank 1995, p. xiii). Several missionary grammars however were still produced

Box 1.9: List of notable works

马若瑟汉语札记

1652 年，卫匡国完成了《中国文法》一书。他将汉语划分出十大词类，并在动词部分对汉语主动语态和被动语态有所介绍。1703 年万济国的《华语官话语法》，依旧沿袭拉丁语法，进行词类划分，并对其形态和范畴进行描写。除此之外，对汉语的一些造句规则（主要是主动句型和被动句型）也进行了归纳说明。1728 年，马若瑟的《汉语札记》首次区分了汉语口语和书面语的语法规则，指出了汉语中"实字"与"虚字"的区别。不过马若瑟并没有依照拉丁语法构造汉语语法规则，而是从语言事实出发对汉语进行描写，这样的做法更符合汉语言的实际情况。

进入 19 世纪，随着欧洲汉学的兴起，越来越多的传教士和专业学者投身于汉语的语法研究中。法国汉学家雷慕莎的《汉文启蒙》在词类划分的基础上，探讨了一些句法规则，并指出汉语的语法不能用西方传统的拉丁语法去生搬硬套。在传教士方面，马礼逊的《通用汉言之法》其系统性比之前的语法专著更强，但主要内容还是集中在词法上，句法部分只归纳了九条规则，仅占五页篇幅。马士曼的《中国言法》在汉语语法本体研究上，也是由八大词类入手着重讨论词法，对句法的讨论较为简单，甚至有时句法与词法的界限并不清晰，将许多构词法放入了句法部分。艾约瑟的《中国官话语法》在前人基础上有所进步。除词法外，他对汉语句法规则的考察也颇为详细，这是此前西方人研究中所少见的。

1.4.3 Modern descriptions

Modern linguistic description of Classical Chinese has a clear starting point: the publication of 《马氏文通》'Ma's ' by Ma Jianzhong (馬建忠).

TODO: 中等国文典, American structuralism in China for the study of Mandarin, Any other grammars in English besides Pulleyblank?

1.5 Texts

The great historical work 《史记》('lit. historical records'), often known as *Records of the Grand Historian* in English (a translation of 太史公记, the earliest known title of the work), laid the paradigm of official historiography of all Chinese dynasties after Han. The author 司马迁 *Sima Qian* is known as the *Lord Grand Historian* or 太史公. 太史 'grand historian' was the title of

Box 1.10: List of Classical texts

- Pre-Qin, and Han
- 唐宋八大家?

1.6 Remarkable features

Classical Chinese has several notable typological features.

Box 1.11: Remarkable features

- Wordhood
- Part of speech
- Topic-comment
- "Coverb", or is there real preposition
- Prosody (and hence a chapter on phonology and writing system)
- The chapter on phonology and writing system can be very hard: lots of historical facts
- Passivization and so on
- Higher tolerance of ad-hoc recategorization: 名作动, 使动意动, etc.

Chapter 2

Grammatical overview

2.1 General principles

Like all natural languages, the syntax of Classical Chinese can be divided into the syntax of the clause (§ 2.2) and the syntax of the noun phrase (§ 2.3), both of which contain a hierarchy of grammatical systems. Concepts like noun-hood and verb-hood can then be defined according to the syntactic environment: a noun is what appears at the center of a noun phrase (NP), and a verb is what appears at the center of a clause (e.g. the distinction between nominal predication and verbal predication in § 2.2.1.3). In this sense, noun-hood and verb-hood in Classical Chinese have nothing inherently different from their counterparts in other languages.

Besides the syntactic constructions, a language also has a *lexicon* that dictates the details of whether and how a root or derived stem or a larger construction appears in certain syntactic environments and be phonologically realized. Here, we have a slightly different definition of parts of speech tags like *noun* or *verb*: they are defined *lexical* labels representing the structure of the *lexicon*, not the *grammar*. The *noun* class now represents a group of lexical items with shared grammatical properties, like frequently appearing as heads of noun phrases and having certain morphological properties. Demarcation of *this* definition of parts of speech often shows considerable cross-linguistic variance as well as language-internal irregularities (for instance, the English adjective *worth* takes an object, a structure not otherwise seen for other adjectives) and is discussed in § 2.4 for Classical Chinese.

Box 2.1: Comparison between Latin and English nouns

For example, to say "the Latin word *canis* is a noun" means to say that the form *canis* usually appear as the head of an NP, that it carries an inherent gender feature and a number feature, and that its inflection pattern follows one of Latin nominal declensions. Modern English does not have rich inflectional morphology but does have nominal modification constructions (e.g. *a* $[dog]_{nominal\ (not\ NP)}$ *tag*), so saying that *dog* is a noun means something different with saying that *canis* is a noun.

We note that *canis* can be further analyzed as a root plus an ending. The Latin word *canis* is actually a bundle of the root *cane-*, the masculine gender, a case feature (here nominative), a number feature (here singular), and the fact that it is the head of some complete NP. On the other hand, the root *cane-* appearing as the main verb of a clause is impossible, because a bundle of the root *cane-* plus some verbal features is *not* in the mental dictionary of a Roman. Nominal attributes are not possible in Latin, again because the mental dictionary of Romans does not contain anything like the root *cane-*

without the head status of a NP.

The Latin form class *noun*, then, means the bundle "a gender feature, a case feature, a number feature, and the head-of-NP status" plus how it is morphophonologically realized (i.e. the five declensions). The English concept of *noun* is quite different from that. Indeed, if we accept the hypothesis that abstract principles of language structures are more or less the same cross-linguistically, then the lexicon *has to* be highly diverse across languages because it is exactly the locus of language variance, besides morphophonology.

Turning back to Classical Chinese, when we say Classical Chinese has a noun-verb distinction in the *lexical* sense, we need to demonstrate that the lexicon of Classical Chinese has two largely non-overlapping groups, the elements of which regularly head noun phrases and clauses and have distinct properties in other morphosyntactic processes, respectively. In § 2.4 we demonstrate that this is indeed the case.

The lexicon is often said to contain *words*, which is inevitably not defined in a clear-cut way (§ 1.3.1), or *lexemes*, which involves the distinction between derivation and inflection. This, again, is not a clear distinction: the relevant parameters include the degree of fossilization (§ 1.3.1; derivation is often understood as creating new words, i.e. fossilized units in the lexicon) and the type of grammatical categories involved (tense, for instance, belongs to inflection, as it's a clause-level category; nominalization of a root on the other hand is considered derivation). The two parameters show a certain degree of consistency, as a smaller grammatical unit also tends to fossilize faster, but notable disagreements exist: a tense-inflected form can be fossilized as well (e.g. English *used to do sth.*), and "small" constructions can be synchronically productive as well. We discuss relevant details in TODO

2.2 The overall clausal structure

Like all other languages, a Classical Chinese clause can be a simple clause or a complex one constructed from subordination (§ 2.2.8) and coordination (§ 2.2.9). A simple Classical Chinese clause can be divided into a nucleus clause (§ 2.2.1, § 2.2.2) plus discourse-related devices, including its speech act (§ 2.2.7) marked by sentence final particles (§ 2.2.2.4), and topicalization or focalization (§ 2.2.6). Topicalization can also happen for a complex clause (§ 2.2.9.2).

It appears that only sentences – clauses that appear as utterance units – have the aforementioned discourse-related devices like topicalization and sentence final particles. All embedded clauses (apart from direct quotations; § 2.2.3.2.1) in Classical Chinese do have these devices.

The nucleus clause may be either a nominal predicate clause (§ 2.2.1) or a verbal clause (§ 2.2.2). Both constructions seem to have a well-defined subject position (which is not the same as the topic), which however is often left blank (§ 2.2.5). Classical Chinese lacks obligatory tense, aspect, mood (TAM) marking but does have a TAM adverb system, and certain semantic subtleties displayed by sentence final particles also suggests grammaticalized TAM information (§ 2.2.4).

Box 2.2: Clause types

In Mei (2015, p. 131), he classifies clauses into 说明句, 描写句, and 叙事句. The classification is comparable to that given in http://area.hcjh.tn.edu.tw/noise/hcjh-ca/4-b.htm#0303 . Mei doesn't mention on which basis he makes this distinction. In the latter source, it seems the distinction is made based on the type of the predicate. Thus a 描写句 is a stative (adjectival?) clause, and a 判断句 is a nominal predicate construction, and a 叙事句 is a verbal predicate construction that is not a 描写句. So what does Mei mean by 说明句? The term appears in Li (2004) as well.

We can go to places where he mentions the term. p. 445: 矣 is for 叙事, and 也 is for 说明. pp. 264-265: 事件句 (叙事句) 和非事件句（描写句和说明句）The distinction is also mentioned in http://paper.wenweipo.com/2018/02/14/ED1802140024.htm

So it's related to the event structure. We need to know where the event structure resides in the vP-TP-CP hierarchy. Particularly, we need to identify *where* the category of this distinction lies. I think probably that's related to the aspect: consider the distinction between a habitual clause and a prototypical "event" clause.

The distinction has syntactic significances. We note that certain topicalization constructions seem to be only compatible with 说明句 (Box 2.16).

Box 2.3: Adjectival predication of a non-finite clause?

Consider 其為人也，發憤忘食，樂以忘憂，不知老之將至云爾: the subject of 發憤忘食，樂以忘憂，不知老之將至云爾 is clearly 其 i.e. Confucius. But then consider the fact that the subject *cannot* appear after 也. This seems to indicate that we are dealing with some sort of subject sharing construction.

Also, consider 为文斐然可观矣: here 斐然客观 semantically modifies 文?

See also Box 2.16.

2.2.1 Nominal predication

2.2.1.1 Real nominal predicates

The top-level structure of a Classical Chinese clause may contain a (optional) subject and a NP acting as the predicate (1, 2). A nominal predicate may express an "is-a" relation between the subject (see § 2.2.5.2 for discussions on the meaning of the term) and the predicate, which is the case of (1). Some nominal clauses however express a possessive relation between the two (2).

(1) [秦]$_{\text{subject}}$， [虎 狼 之 国]$_{\text{predicate}}$
Qin tiger wolf GEN country
'Qin is a country of tigers and wolves (i.e. cruel and not reliable).'

(2) [蟹]$_{\text{subject}}$ [六 跪 而 二 螯]$_{\text{predicate}}$
crab six leg CONJ two claw
'A crab has six legs and two claws.'

Box 2.4: The possessive nominal predicate construction

It seems the predicate in the possessive nominal predicate construction can never be a bare noun without any modification. The modification can be a numeral or an attributive.

(3) 王六军，大国三军

(4) 秦王 [为人]$_{\text{Box 2.16}}$，蜂准，长目，挚鸟膺，豺声，少恩而虎狼心

Another problem is that the 者-也 construction seems to be incompatible with the possessive nominal predicate.

Negation in Classical Chinese nominal clauses is usually expressed by 非, placed before the nominal predicate (5).

(5) 凡群臣之言事秦者，皆奸人，非忠臣也

The nominal predicate is compatible with TAM adverbs

(6) 且藺相如素贱人

2.2.1.2 Topicalization of nominal predicate construction

(1) is much less frequent than the 者…也 construction, often known as 判断句 in Modern Chinese or the 'judgemental clause'. A judgemental clause usually contains a particle 也 (§ 2.2.2.4) at its end (7), or a particle 者 after the subject (8), or both. It seems that the judgemental clause is better analyzed as a topic-comment construction (§ 2.2.6.1).

(7) [城 北 徐-公]$_{\text{topic: NP}}$，[齐-国 之 美-丽
city north NAME-GONG Qi-country GEN beautiful-beautiful
者]$_{\text{comment: § 2.3.2}}$ 也
REL SFP
'Mr. Xu from the north of the city is a handsome guy in the country Qi.'

(8) [兵]$_{\text{topic: NP}}$ 者， [不 祥 之 器]$_{\text{comment, predicate}}$
weapon TOPIC NEG fortunate GEN instrument
'Weapons are not auspicious.'

2.2.1.3 Distinction between a nominal clause and a verbal clause

Note that the term *nominal* in *nominal predication* or *nominal clause* refers to the fact that the predicate is structurally a NP, not whether the head of the predicate usually appears like a noun or a verb in a dictionary (§ 2.1). In some sentences although the predicate of a clause mostly appears as the head of a NP and therefore may be referred to as a noun in dictionaries, the clause is clearly a verbal clause because it expresses a dynamic event and not just a state, the possibility of TAM markers, etc., as in (9). Here 水 'water' is used as a verb, meaning 'swim', which is also modified by the modality auxiliary 能 'can'.

(9) [假 舟 楫 者]$_{\text{subject}}$，[非 能 水 也]$_{\text{predicate: VP}}$……
draw.help boat paddle REL NEG can swim SFP
'Those who draw help from boats and paddles cannot swim, ...'

There are cases where the meaning of the predicate is comparable to that of a real nominal predicate. We still classify them as verbal clauses, because of their similarity with prototypical verbal clauses with respect to negation, TAM modification, TODO

(10) 大楚兴，陈胜王

(11) 然而不王者，未之有也

On the other hand, there is one thing a nominal predicate can do while a verbal predicate *cannot* do: a nominal predicate can be topicalized (§ 2.2.6.1, 61).

2.2.1.4 Copula constructions

All the constructions mentioned above are without a copula. In the pre-Classical copula age there is a copula 惟, which however had largely died out of use in Classical texts. Meanwhile, grammaticalization had added several copulas to Classical Chinese (Pulleyblank 1995, pp. 20-22).

2.2.2 Verbal predication

The structure of clauses with verbal predicates is much more complicated, and the details can only be described in the following sections. In this section we overview grammatical systems within verbal clauses.

2.2.2.1 Constituents and ordering

In clauses with verbal predicates, the constituent order of core constituents of transitive clauses is almost always SVO (12, 13). Intransitive clauses have a SV constituent order (14). The usage of the term *subject* is justified in § 2.2.5, and the contents of a verbal clause besides the subject is often defined as the verb phrase (VP). VPs can be coordinated (§ 2.2.9). Prepositional complements are also placed after the verb (15). The term *object*, without specification, means any argument in the VP that is not marked by a preposition (§ 2.2.3.1.2).

(12) [子张]$_{\text{subject (§ 2.2.5)}}$ [[学]$_{\text{verb}}$ [干禄]$_{\text{object}}$]$_{\text{predicate: VP}}$

(13) [子]$_{\text{subject}}$ [奚]$_{\text{reason (§ 2.2.3.6)}}$ 不 [为]$_{\text{verb}}$ [政]$_{\text{object}}$

(14) 君子不器

(15) 君子博学于文

Object pronouns however can be extracted before the verb in negative clauses (16), leading to a SOV order. An interrogative object pronoun can also be fronted (17).

(16) 恐 [年岁之 [不吾与]$_{\text{VP: Neg-OV}}$]$_{\text{complement clause}}$

(17) 以五十步笑百步，则 [何如]$_{\text{SOV interrogative clause}}$

2.2.2.2 The structure of the verb

It is possible that the main verb of a verbal clause contains more than one root. Such a verb is known as a complex predicate.

> **Box 2.5: Classical Chinese complex predicate**
>
> Directional complement and resultative complement

2.2.2.3 Positions of modifiers

Adverbial constituents in the nucleus can be divided into TAM ones and so-called peripheral arguments, including location, manner, instrument, etc. The peripheral arguments can be post-verbal (18, 19, 20) or pre-verbal (21, 22, 23), with the pre-verbal order gaining popularity as time went by. The linear order of peripheral arguments is similar to that in Mandarin (He 2005, pp. 286-287).

(18) 侍饮于长者
(19) 孟孙问孝于我
(20) 祷尔于上下神祇
(21) 韩生南向坐
(22) 於人之罪无所忘
(23) 为人谋而不忠乎

The TAM adverbials are almost always preverbal.

(24) 文王既没，文不在兹乎
(25) 孔子既得合葬于防
(26) 我未之能易也

When TAM adverbs and peripheral arguments both appear before the verb, the order is always TAM > peripheral argument. The reverse order is never attested. The whole VP therefore can be analyzed as a core VP plus peripheral arguments surrounding it, plus TAM adverbs preceding the pre-verbal peripheral arguments. The clause then is the complete VP plus the subject.

(27) 三王 [既]$_{\text{TAM}}$ [以]$_{\text{instrument}}$ [定法度]$_{\text{VO}}$

Box 2.6: Adverbials combination

Is it possible to use multiple pre-verbal peripheral adverbials? What's the relevant order constraint?

Box 2.7: Position of adverbials in SOV case

Where to place adverbials in SOV case?

Box 2.8: Position of negator

Where is the position of the negator?

2.2.2.4 Sentence final particles

Classical sentence final particles have a variety of functions. It may mark the interrogative force (28), a judgemental meaning (29), and aspectual values (30).

(28) 大车无輗，小车无軏，其何以行之哉
(29) 人而无信，不知其可也
(30) 温故而知新，可以为师矣

It seems a sentence final particle can be shared by two conjuncts.

(31) 虎者，戾虫；人者，甘饵也

In early texts, a sentence final particle can be inserted after the main verb.

(32) 巧言令色，鲜矣仁

2.2.2.5 The gerundive construction

Classical Chinese has one gerundive construction, which may appear as the object (33) or as a subordinated clause, like a conditional or temporal clause (34; § 2.2.8). The construction is sometimes known as nominalization. We reject this analysis, because it seems the construction does not admit adjectival modification, while there is nothing looking like an object (e.g. 子 in 34) in prototypical Classical Chinese noun phrases. The construction therefore has a structure comparable to the English gerundive non-finite clause *his playing national anthem* and is structurally different from the noun phrase.

(33) 王 如知 此，则 无 望 [民 之 多 于 邻
king if know this then NEG hope people GEN more than neighbor
国]$_{\text{object: gerundive}}$ 也
country SFP
'If Your Majesty knows this, then don't expect your people to be more plentiful than your neighboring countries' people.'

(34) [父母之爱子]$_{\text{condition: gerundive}}$，则为之计深远

2.2.3 Argument structures

2.2.3.1 Core argument structures

2.2.3.1.1 DO, BE and BECOME Consistent with cross-linguistic generalizations, in Classical Chinese, a verbal clause can be about an intentionally initiated event (DO; § 7.1.1), a state (BE) or a change of the state (BECOME; § 7.1.2). The DO type can further be divided into the transitive and intransitive classes. BE and BECOME clauses are intransitive by definition. The distinction between the three classes has consequences for animacy and volition of the subject (see the sections referred above) as well as the viability of certain grammatical processes (§ 7.1.1.1).

BECOME and BE clauses are often inputs to the synthetic causative construction, resulting in CAUSE-BECOME/BE clauses (§ 2.2.3.4.1), which often develop lexicalized usages (§ 2.2.3.5).

2.2.3.1.2 Prepositional arguments and applicative constructions Prepositional arguments can also be observed in Classical Chinese. In (35), for example, the prepositional phrase 于车 is the *source* of the event. A prepositional argument can also appear in a transitive construction, coding a wide varieties of semantic roles, like the recipient (36, 37), or the target of a question (38). It is not possible for a prepositional argument to appear before the object. Classical Chinese also does not have quirky subjects: it is not possible for a prepositional phrase to appear in the subject position.

(35) 公 惧， 队 于 车
king afraid fall from carriage
'The king was afraid and fell from the carriage.'

(36) 成王、康王……故赐之以重祭

(37) 秦复予我河外及封陵为和

(38) 季康子问政于孔子

Classical Chinese has applicative constructions that turn an argument structure containing a prepositional argument into a double object construction (39). In this case, the argument corresponding to the prepositional argument in (38) behaves like the monotransitive object in constituent orders and in valency decreasing (§ 2.2.5.1.2, § 7.3.1.3), which means it is somehow more "external" or "subject-like" (§ 2.2.5.1.2).

(39) 上问上林尉诸禽兽簿

It should be noted that there exists another type of double object construction derived from verbs with a prepositional argument: the preposition of the prepositional argument may be omitted after the verb: compare the prepositional (40) and the double-object example (41). The omission seems to be in line with the omission of the preposition in post-verbal locative arguments (cf. 56).

(40) 有献不死之药于荆王者

(41) 请献盆缶秦王

In some languages, the argument structure of verbs meaning giving and receiving seems to contain a small clause. In English, for instance, we have *give this to him and that to her*, and in Latin we even have standalone small clauses like *Deo gratias*. No trace of such possessive or directional small clauses is found in Classical Chinese: double object clauses with giving or receiving meanings seem to be analyzable as applicative clauses (Mei 2015, pp. 416-421).

2.2.3.2 Verbal complementation

Verbal complementation in Classical Chinese includes various complement clause constructions, and clauses with sub-clausal complements.

(42) 子使漆雕開仕

(43) 雍也可使南面

2.2.3.2.1 Direct quotations Direct quotations in Classical Chinese may appear within the VP just like an ordinary object (44). Note that the quoted content is a *sentence*, which includes a sentence final particle, which is usually not allowed in other embedded clauses (TODO: ref). We however note that the verb in the V-quotation construction illustrated in (44) is limited to 曰. For other "speaking" verbs, the direct quotation is introduced in the way of (45).

(45) can be analyzed as an object sharing construction, where 问于子贡 and 曰 are coordinated at the argument structure level (§ 2.2.9.1), and the direction quotation is an argument of both 问 (cf. 38, where the direct quotation in 45 is replaced by the NP 政 'politics'). We however note that 曰 has possibly grammaticalized in Classical

texts. In (46), the constituent introduced by 曰 is a proper name, and clauses with 曰 as the main verb where the subject is a person and the object is a proper name are rare if not impossible. If 曰 in (46) is understood as a marker of a direct quotation, however, the sentence makes sense: the second argument has a *metalinguistic* usage, whose semantic interpretation is the quoted 灵台 itself, without referring to anything in the real world.

(44) 子曰：“不患人之不己知，患不知人也。”

(45) [子禽]$_{\text{subject: NP}}$ [问]$_{\text{verb}}$ [于 子贡]$_{\text{target: PP}}$ [曰：“…”]$_{\text{direct quotation}}$
NAME ask at NAME say
‘Ziqin asked Zigong: …’

(46) [谓]$_{\text{verb}}$ [其台]$_{\text{object: NP}}$ [曰 [灵台]$_{\text{proper name}}$]$_{??}$

Therefore, in Classical Chinese, 曰 is both used as a lexical verb (44) and a grammaticalized marker of direct quotations. What is quoted can be a sentence (45) or a noun phrase (46). Under this analysis, examples like (47) are probably clauses with nominal predicates, with the predicate being a direct quotation.

(47) 其名曰觚

2.2.3.2.2 “Prototypical” complement clause constructions

(48) 臣窃以为 [不便於君]

(49) [知和而和，不以礼节之]，亦不可行也

2.2.3.2.3 Pivot constructions or argument sharing Certain verbs have two internal arguments, the first is the object of the main clause, and the second is a complement clause, whose subject is the aforementioned first object, i.e. the object of the matrix clause. This construction is sometimes known as the *pivot construction* (Pulleyblank 1995, p. 40) or 兼语式 in Chinese (Mei 2015, p. 375).

A clear instance of the pivot construction is the analytic causative construction (§ 7.2.3). (50) is an example. Uncontroversial pivot constructions are limited in number in early texts, because many of them can also be analyzed as argument structure-level coordination (Mei 2015, p. 376; § 2.2.9.1).

(50) 令军勿敢犯

> **Box 2.9: Other verb frames**
>
> Control construction, etc.

2.2.3.3 Possession in argument structure

(51) 陈胜者，阳城人也，字涉

2.2.3.4 Valency alternation

In a language with frequent information packaging operations, whether what is intuitively perceived as valency alternation is truly valency alternation and not just information packaging needs discussions. We discuss this topic in e.g. § 7.3.1.1.

2.2.3.4.1 Valency increasing Various valency increasing constructions exist in Classical Chinese, which all append a subject, i.e. an external argument (§ 2.2.5.1) to an existing argument structure.

The most productive valency increasing construction is probably the synthetic causative, whose outputs are "transitive" CAUSE clauses (§ 7.2) that are similar to but subtly different from transitive DO ones.

Classical Chinese seems to already have a prototype of what is later known as the disposal construction or the 把 construction in later Sinitic languages.

(52) 尽以其宝器赂献于周厘王

2.2.3.4.2 Valency decreasing Classical Chinese also has valency decreasing constructions, which *suppress* the subject and promote an internal argument (§ 2.2.5.1.2) to the subject position (§ 7.3). Valency increasing after valency decreasing is also possible: (53) is an example of a causative clause based on the pseudo-passive construction.

(53) ……杀御叔 (=9 in § 7.2.1)

2.2.3.4.3 Applicatives In § 2.2.3.1.2, we see that Classical Chinese also has productive applicative constructions, and the resulting argument structure is subject to valency decreasing.

2.2.3.5 Argument structure and verb classes

The only fundamental constraint to whether a stem appears in a transitive DO or intransitive DO or BECOME construction is its semantics. In world languages, however, whether a root or a stem is compatible with a certain verb frame is dictated by the lexicon of the language, and a group of root-environment complexes with shared properties is known as a part of speech (§ 2.1). Thus we can say if a *verb* (and not the clause it heads) is transitive or intransitive, or whether it is an action verb (i.e. a DO verb), a stative verb (i.e. a BE verb), a internally caused change of state verb (i.e. a BECOME verb) or an externally caused change of state verb (i.e. a CAUSE verb, which may be CAUSE-BECOME or CAUSE-BE or even without an intransitive counterpart). Analysis of argument and event structures is closely related to verb classification.

2.2.3.6 High-level categories of the VP

Certain constituents within the VP can appear before the main verb. In (54), the preverbal constituent 以其妹 seems to be a fronted prepositional argument. It cannot be the sentential focus, because the subject 季康子 seems to stay in-situ with no pause after it. It is likely that a VP-internal focus position exists in Classical Chinese, which, cross-linguistically, is not rare (e.g. see Danckaert (2011)). A further piece of evidence suggesting the in-VP analysis of (54) is that the fronted prepositional phrase may further undergo preposition-object inversion (55), an operation that is otherwise not observed in Classical texts and also not motivated: the best explanation of this inversion seems to be focalization (Mei 2015, p. 323).

(54) 季康子 [以其妹]$_{\text{VP-focus: prepositional phrase}_i}$ 妻之 $-_i$

(55) 室於怒而市於色

It is possible to omit the preposition of certain post-verbal

(56) [大-王]$_{\text{subject: NP}}$ 见 [臣]$_{\text{object: NP}}$ [列 观]$_{\text{locative: NP}}$
great-king see servant/minister regular palace
'Your Majesty see (your) servant (i.e. me) at a regular palace.'

2.2.4 Tense, aspect, modality, and things like that

The *meaning* of tense, aspect, and modality can be expressed by various syntactic devices, not all of which should be considered *syntactic* TAM categories comparable to, say, English *should have been doing*. *Syntactic* TAM categories are within a single clause and form a system with implications for e.g. linear orders of auxiliaries and adverbs. Thus *should*, *have*, and *been* in *should have been doing* are considered real auxiliaries and their existence demonstrate that English has a syntactic TAM system, as their relative orders cannot be shifted even with semantic motivations. English adverbs like *now* or *frankly* are also likely a part of the TAM system because they have to follow the same kind of order constraints (Cinque 1999). On the other hand, English *be able to* is not a grammaticalized TAM marker (yet), as *many of us would have been able to have made this kind of bomb* is attested, where the PERFECT marking appears both before and after *able*, modifying both the fact many of us being able to do something and the thing being done (i.e. making this kind of bomb), strongly suggesting that we are looking at a biclausal construction.

An analysis of time in Classical Chinese along this line however proves hard, because we do not have native speakers and certain grammatical details may never be known for certain: it is not possible to know that a certain construction is not possible, and therefore there is always room to argue that a certain word is not fully grammaticalized. The analyses below therefore are all tentative to a certain degree.

In Classical Chinese the main markers that represent a grammaticalized TAM system seem to be time adverbs, which follow a linear order like epistemic modality> tense, as is mentioned above for English (§ 8.1). Certain

Box 2.10: In search of auxiliaries?

See e.g. here.

- 克”“能”“堪”“可”“可以”“可得”“得”“足”“足以”等
- 欲”“肯”“將”“宁”“敢”“忍”“愿”“屑”“慭（yìn）”等，其中“敢”“忍”“屑”通常用于否定句，“肯”“（不）敢”“（不）忍”“（不）屑”沿用至今。“欲”“將”“愿”“慭”表示主观的希望或打算，可以翻译成“希望”或“打算”。例如：
- 当”“如”“宜”“任”“合”“应”，其中“合”“应”
-

能 does not seem to be a prototypical auxiliary. The main argument against its status as an authentic auxiliary is the prevalence of sequences like 不能不 (as in 人主不能不有遊觀安燕之時). Here the two negations means 'not doing ...is impossible', and the structure is routinely used to emphasize the core VP. The structure of these examples seems to be better analyzed as a complement clause construction, as nowhere else in Classical Chinese can we find two same negators in a single clause. (非不 is possible, but not 不不)

cf. Negation in Cartography

2.2.4.1 Speaker-oriented categories

The most "high-level" TAM categories are usually categories related to the speaker's evaluation of the situation being talked about. TAM categories falling under this type include evaluative adverbials (like *frankly* at the start of a sentence), evidentiality, and epistemic modality.

Classical Chinese has epistemic

2.2.4.2 Tense

The existence of a tense system can be demonstrated by the subtle semantic difference caused by the sentence final particle 矣, although the particle itself is likely not a dedicated tense marker (§ 8.3.1). The past and the future have adverb markers (§ 8.1.2).

2.2.4.3 Viewpoint aspect

2.2.4.4 Lexical aspects

Whether a clause is a DO clause or a BECOME clause or a BE clause is also related to the lexical aspect of the clause, which in turn may have non-trivial interactions with TAM categories.

> **Box 2.11: Aspects of TAM marking**
>
> Are TAM markers allowed in the pivot construction? This is related to whether Classical Chinese has infinitives (Mei 2015, p. 375).

2.2.5 Subjecthood

In § 2.2.1.1 and in § 2.2.2.1, we both mention the concept of *subject*, which needs justification. Further, the bipartite division of a verbal nucleus into a subject and a VP in § 2.2.2.1 means that the subject is in some senses *external*, while other arguments within the VP are *internal*. Cross-linguistically, it is not impossible for a language to demonstrate two types of *externality*, one based on argument structure properties like obligatory argument omission in control constructions or binding of reflexive pronouns (§ 2.2.5.1), another based on clausal pivot properties like subject sharing in coordination and relativization (§ 2.2.5.2). The two being different means syntactic ergativity, which is rare among attested world languages (Aldridge 2008), and is absent in Classical Chinese.

Another problem is the distinction between subject and topic. Since both the topic and the subject appear at the beginning of a clause, the distinction between the two seems unclear. We can even go as far as claiming that Classical Chinese has only information structure and no argument structure in its syntax (Mei 2015, p. 122). The matter is further complicated by the fact that Classical Chinese has no native speakers now and detailed grammaticality tests are not available, and that Classical Chinese is a pro-drop language so obligatoriness is not a viable criterion. Still, we believe that the existing evidence is sufficient to justify postulating a subject grammatical function in Classical Chinese besides the topic position, which is for information structure marking (§ 2.2.5.3).

2.2.5.1 Subjecthood in argument structure

2.2.5.1.1 Obligatory relation between semantic roles and linear order We observe that the semantic relation between some clause-initial NPs and the verb is fixed by properties of the verb, while the semantic relation between some clause-initial NPs and the verb is more flexible, and these NPs are related to some internal positions of the nucleus clause. We therefore rightfully call the first type of clause-initial NPs subjects, and the second type of clause-initial NPs topics.

(57), for example, contains three constituents, and therefore can only be a verbal clause, with the last constituent being the object. The verb 客 is a derivation from the noun 客. Such a derivation, according to our experiences with other Classical Chinese texts, can only be causative or tropative or benefactive. By considering the context we will know the clause is a tropative one and the right translation is 'Lord Mengchang considers me as a guest.' Therefore, by virtue of being at the initial of the clause, the NP 孟尝君 *has to* be understood as what initiates the event, can be neither the patient nor peripheral roles in the event (e.g. an instrument). The *obligatory* relation between the verb and the NP 孟尝君 clearly shows the latter is a subject, and not a topic.

(57) 孟尝君客我

It should be noted that the subject can be the patient, as in (58). This however again is an *obligatory* semantic relation between the verb 定 and the NPs 国 and 天下. By virtue of being the only argument of 定 and appearing before the verb, 国 and 天下 are obligatorily understood as the patients. They cannot be understood as, say, the location of the event ('*Someone causes peace (i.e. 定) to something else in the state (国) and the universe (天下)'). Clauses like (58) are therefore better analyzed as valency alternation constructions.

(58) 国定而天下定

2.2.5.1.2 Subjecthood in argument structure and valency alternation The existence of a subject on the level of argument structure is relevant in valency alternation, as in e.g. § 7.1.3 and § 7.2.1: a structure with a subject is too "big" for certain operations.

An internal argument – always an object, not a prepositional argument – in a clause may be promoted to the external subject position in another clause (e.g. § 7.1.2). It is possible for multiple objects to co-exist, and as all double object constructions in Classical Chinese seem to be related to one applicative construction or another (§ 2.2.3.1.2), the rule is that the object introduced by the applicative gets promoted to the subject position. We may say that the object created by the applicative is the "second most external" argument. This hierarchy of externality of arguments is not uncommon in world languages (Box 2.12).

> **Box 2.12: Multiple external arguments?**
>
> In Japhug, it is possible to have a *causer→instrument→agent→patient* argument structure, and the personal indexation marker seems to be decided by first taking the two most internal arguments and decide which is more salient on the empathy hierarchy, and then compare the result with the third most internal argument and decide which is more salient, and finally compare the result of the last step with the most external argument; hence 1→3→2→3 is morphologically equivalent to 1→2 (Jacques 2021, p. 310, p. 584, (116), p. 848, (67)).

In Classical Chinese, the synthetic causative construction generally cannot be applied to an argument structure already with a subject (§ 7.2.1), so structures like this are not possible. Yet as is seen above, a similar hierarchy can be built by the applicative.

2.2.5.2 Subject as clausal pivot

In § 2.2.5.1, we see that subjecthood can be defined in the argument structure in verbal clauses. Yet properties commonly attributed to subjecthood are not just about the argument structure. For instance, in the nominal predicate construction (§ 2.2.1), we call the first NP the subject, and there is no such thing as the argument structure there. What we want to know is whether both verbal and nominal clauses in Classical Chinese have a pivotal position in it which everything else "revolves around" which could be called the *subject*.

The most clear criterion that defines clausal pivotal subjecthood is probably coordination: if when clauses are coordinated, one constituent seems to be shared by all of them, then this constituent is probably the clausal pivot. It turns out that what is defined as the subject according to its behaviors in the argument structure indeed is also the pivot in coordination (§ 2.2.9, 75). Note that two coordinated clauses can also share a topic, but there are signs which tell us that what is shared is the topic and not the subject (e.g. § 2.2.9, 76, where the topic is the object of the first clause and the subject of the second clause).

Box 2.13: Definition of VP

Can a verbal and a nominal predicate be coordinated?

Box 2.14: Subject and TAM

Subject and TAM in English are closely related. (e.g. control construction) What about Classical Chinese?

The observation that the argument structure subject in verbal clause turns out to be the clausal pivot and that the TODO: nominal clause justify the usage of the term *subject* outlined in the beginning of this section.

2.2.5.3 Comparison with topic

We have already argued that in every Classical Chinese clause, there is a (possibly empty) subject position, which is largely *independent* to information structure factors and therefore is not a topic. On the other hand, authentic, information structure-related topics are marked by devices not always available for subjects (like the particle 者 or a pause; § 2.2.6.1). So indeed subject and topic are two distinct concepts in Classical Chinese.

This does not mean that there are no blurry cases. This probably leads to some scholars to treat any constituent that seems to be "external" as a subject (e.g. Li 2004, p. 41), and hence the topic is a "big subject" (Li 2004, p. 42).

2.2.6 Information packaging

2.2.6.1 Topicalization

Topicalization in Classical Chinese is usually marked by adding the particle 者 after the topic. In the reading tradition, a pause is often inserted after 者, which cross-linguistically suggests topicalization (59, 60). In these examples, the subject is topicalized. We note that the structure of these two examples is comparable to that of the "judgemental clause" (§ 2.2.1), which obliges us to analyze the judgemental clause as topicalization of the nominal predicate construction.

(59) 此二人者，实弑寡君

(60) 单父人吕公……吕公者，好相人，……

What is topicalized is of course not restricted to the subject. This fact is a piece of evidence supporting the distinction between subject and topic in Classical Chinese. In (61), the comment clearly has a nominal predicate. What is promoted to the topic position however is not the subject of the nominal predicate construction, but the predicate. The subject is likely the *focus* and not the topic (Mei 2015, p. 138). Topicalization of other clausal constituents is also possible (e.g. § 2.2.9, 76).

(61) $[\text{仁之实}]_{\text{topic: NP}_i}$，$[[\text{事亲}]_{\text{subject: ?}}\ [\text{是}]_{\text{predicate: pronoun}_i}]_{\text{comment}}$ 也

We also note that the 者…也…framework is not limited to topicalization of nominal predicate clauses (i.e. "judgemental clauses"). For instance, we have (62), in which the sole argument in a existential clause is topicalized, and the comment receives 也 as its sentence final particle.

(62) 然而不王者，未之有也

We also note that topicalization can happen multiple times (63).

(63) 万乘之国，弑其君者，必千乘之家

Box 2.15: Dangling topic

Are there dangling topics in Classical Chinese? If not, it's a another piece of evidence supporting the distinction between subject and topic.

Box 2.16: A 之于 B 也

(64) 寡人之于国也，尽心焉耳矣

The original structure is probably 寡人于国尽心. The fact that 寡人 and 于国 are subtracted from the nucleus clause is rather unusual. And similar examples exist. Consider 史记夫乐之与音，相近而不同. The original structure is 乐与音，相近而不同, and the insertion of 之 is rather strange.

(65) 其为人也，发愤忘食，乐以忘忧，不知老之將至云尔

A 之于 B 也 (or A 之为 B 也)，predicate, or A 为 B 也, predicate. The structure seems to be parallel to the English *I, as a concerned citizen, want to emphasize that …*, where *as a concerned citizen* obligatorily modifies the subject *I*. 其为人也 here seems to be a *frame*, somehow comparable to the "global" temporal or locational phrase. Another

issue is that the sentence seems to be unable to represent a specific event: * 昨日，孔子为人也，发愤忘食, while we have *yesterday, as a concerned citizen, I …*

(66) 水之积也不厚，则其负大舟也无力。

2.2.6.2 Focalization

Topicalization is marked by fronting and a pause, but what is fronted and before a pause is not necessarily a topic. In (67), for instance, the verbal predicate is fronted, which likely is not topical. Note that the sentence final particle is fronted as well.

Box 2.17: Fronted SFP

The phenomenon can be analyzed in multiple ways. We may assume that (67) is essentially some sort of cleft construction, in which the subject is first separated from the rest of the clause and then the rest of the clause is focalized. Or we can analyze 矣 as a TAM marker, and not a marker from the CP layer. Or maybe we can argue that markers from CP layers are morphologically verbal and have to be attached to either the main verb or the verb phrase. Which analysis works best depends on whether they are consistent with other phenomena.

- Can we prove that the SFPs are very "high-level" and are above the topic layer? For example, can two clauses with different subjects share a SFP?
- Semantically do 矣 carry TAM meanings?

(67) [[甚]$_{VP_i}$ 矣]$_{focus}$， [[汝 之 不 惠]$_{subject:\ gerundive\ (§\ 2.2.2.5)}$
extreme SFP 2 GEN NEG smart
—$_{predicate_i}$]$_{nucleus\ clause}$

'You are so stupid! (lit. So extreme is your being unintelligent!)'

Box 2.18: A complete overview of the left periphery

See https://referenceworks.brill.com/display/entries/ECLO/COM-000248.xml

2.2.7 Speech acts

The fact that sentence final particles usually appear in sentences and direct quotations and not in other types of embedded clauses strongly indicates that these particles mark discourse-related categories.

2.2.7.1 Sentential aspect

Classical Chinese seems to have a *sentential aspect*, which is a discourse device suggesting the listener that a piece of new information is coming. This category is marked by the particle 矣, whose function is close to the sentence final 了 in Modern Standard Mandarin (Paul 2014b; Pan 2021).

2.2.7.2 Interrogative, exclamative, and imperative

The interrogative speech act, for example, is marked by 乎 and other particles (68). The exclamative speech act is similarly marked by sentence final particles (69). We note that in (69), the sentential aspect marker 矣 appears before the exclamative 夫, which means that the two systems of particles can coexist. The reverse order * 夫矣 is not attested.

(68) 其能久乎？

(69) 吾死矣夫!

2.2.8 Subordination

The term *subordination* sometimes means all kinds of clause embedding. In this section we primarily focus on bipartite clauses with the structure and meaning of 'if ...then ...' or 'when ...', and leave relative clauses and complement clauses to TODO: ref

An overview of subordination constructions in Classical Chinese can be found in Mei (2015, Ch. 3). In all Classical Chinese conditional constructions, the condition usually appears before the consequence (70, 71, 72). The consequence can be marked by 则 (70). Sometimes the marker 则 is dropped (71) but putting it back should never render a sentence ungrammatical (Mei 2015, p. 86). The marker 乃 is also available as a marker of the consequence clause (Mei 2015, p. 87).

The condition clause can also be marked. Classical Chinese distinguishes between realis and irrealis conditional constructions: the former are marked by e.g. 既 (70), while the latter are marked by e.g. 若 (72). This distinction is relevant to the licensing of TAM markers (Mei 2015, p. 81). Other markers for the condition clause are also available (Mei 2015, Ch. 3). We note that the marker 若 is able to appear *after* the subject of the condition clause (Mei 2015, p. 94).

(70) [既来之]$_{\text{condition}}$，[则安之]$_{\text{consequence}}$

(71) [杀女]$_{\text{condition}}$，[我伐之]$_{\text{consequence}}$

(72) [若已食] 则退

> **Box 2.19: Position of condition marker**
>
> When the subjects of the two clauses are shared, it seems 若 obligatorily appears after the subject of the first clause. A possible analysis is to assume that the subordination construction is working at the level of VPs.

An interesting phenomenon is that the condition (73) or temporal clause (73) can be a gerundive one (§ 2.2.2.5). This is not surprising cross-linguistically, as the condition clause or the temporal clause is usually the "subordinate" clause, while the consequence clause is the "main" clause, and it is not uncommon for the subordinate clause in a clause subordination construction to have a non-finite structure. This is observed in for example Japanese and Turkish. Note that the marker 若 can be attached to the gerundive condition clause as well (Mei 2015, p. 98). In some condition clauses, the marker 而, instead of the otherwise genitive marker 之, appears between the subject and the predicate, forming a clause type that is not gerundive and only appears as an irrealis condition clause (Mei 2015, pp. 100-102).

(73) 我之不德，民將弃我

(74) 臣之壯也，犹不如人

2.2.9 Coordination

Explicit marking of coordination is primarily done by the marker 而. When used as a conjunction marker, 而 can be used to link two clauses or two verb phrases with a shared subject (75), but not two nominal constituents. Note that the functionalities of 而 is not restricted to conjunction (Mei 2015, p. 183).

(75) 声伯四日不食以待之，食使者，而后食

2.2.9.1 Coordination of VPs

Recall that a VP contains an argument structure (§ 2.2.3) and a set of TAM markers (§ 2.2.4). Therefore, coordination of two VPs actually has two structural possibilities: coordination of two argument structures, resulting in a *single* situation (and the clause is *not* a prototypical compound clause), or coordination of two full VPs (Mei 2015, pp. 192-201). In languages with TAM inflections, in the first scenario, it is likely that the two verbs have one TAM marker in total, or obligatorily have two identical TAM markers. The distinction may also influence relativization (Mei 2015, p. 207).

Classical Chinese does not have TAM-based verbal inflection, but the distinction between the two can still be told

2.2.9.2 Topic chains as syntactic coordination

An interesting question is the interaction between topicalization and coordination. Mei (2015, p. 217) contends that "topic chains", i.e. several clauses with a shared topic (Mei 2015, Ch. 4 § 3.3), are discourse structures and not syntactic structures. Therefore topicalization happens first, and coordination happens then: after that no further topicalization is possible. He further argues that clauses in a topic chain cannot be linked together by 而. (76) however seems to be a counterexample. This example clearly contains two coordinated clauses. In the first clause 取之于蓝, 之, appearing after the verb, can only be a pronoun, and the only sensible reading of the clause is that 之 (the object) is coreferential with 青 at the initial of the sentence, and 取之于蓝 then means '(people) extract it (i.e. indigo dye) from *Indigofera*.' Therefore, 青 at the initial of the sentence is the object of the first clause and the subject of the second clause, meaning it cannot be the shared subject. This, together with the traditional pause after the first 青, means the first 青 likely is a topic, which means here topicalization happens *after* coordination.

(76) [青]$_{\text{topic: NP}_i}$，$-_i$ 取 [之]$_{\text{object: Pronoun}_i}$ 于 蓝 而 $-_i$ 青 于
indigo.dye pick it from *Indigofera* CONJ blue than
蓝
Indigofera
'Indigo dye, people extract it from *Indigofera*, but it's bluer than *Indigofera*.'

2.3 The noun phrase

The Classical Chinese NP can be roughly divided into the determiner region and the "core" region, the latter known in Huddleston and Pullum (2002) as the *nominal*.[1] The latter is just the head noun plus possible complements and modifications, and the first can be left empty or be a demonstrative, or a "possessor", the role of the latter being not confined to a semantic possessor (Pulleyblank 1995, p. 61). When the "possessor" is present, the particle 之 appears between the possessor and the nominal region (77, 78). When only the demonstrative is present, no marking is present (79).

> **Box 2.20: Determiner region**
>
> Give a comprehensive list of determiners.

(77) 王之诸臣

(78) 马之死者

(79) [此心] 之所以合于王者

2.3.1 Structural template

2.3.2 The nominal region

Pre-head attributives

> **Box 2.21: Pre-head attributive**
>
> Is the following paragraph right?
>
> An interesting feature of Classical Chinese is that adjectives before the head noun seem strongly discouraged. The meaning of, say, 'an ugly big old bear', is canonically expressed by several strategies. One is the 者 construction introduced below, which can be described as a relative clause construction (but with caveats) and seems to have no complexity constraints (81). Semantically non-restrictive attributives can always replaced by clausal coordination.
>
> Multiple adjectives are indeed possible.

The marker 者 and the relative clause construction The marker 者 looks like a relativizer. It is different from relativizers in many other languages in that further structural add-ons can be applied to the fused relative clause formed by it, while the fused relative clause constructions in many other languages are unable to undergo further modification. This seems to be the only productive way to form complex nominals (81).

(80) 马之千里者

(81) 若 [至力农畜，工虞商贾，为权利以成富，大者倾郡，中者倾县，下者倾乡里者]，不可胜数

> **Box 2.22: Relative clause complexity**
>
> Can a relative clause contain a NP that in turn contains a relative clause?

[1] In this note, when the term *nominal* is used as a noun, it refers to the determined region in NPs, while when it is used as an adjective, it refers to the status of being the head of a NP.

Box 2.23: *zhi-zhe* construction

The structure of the 之-者 construction may cause some debates. It can be analyzed as a possessive construction on top of a fused relative clause construction and translated word-to-word into English as '[those who go one thousand miles] of horses'. An interesting question then is whether we have any other appearances of the N 之 V 者 construction where the relation between N and [V 者] is prototypically possessive. It seems this is indeed possible: 城北徐公，齐国之美丽者也.

Under this analysis, 楚人有吹箫于市者 is composed by applying the external possessive construction to 楚人之吹箫于市者

One fact (or is it really a fact?) supporting the determinative analysis of 之-者 is the construction seems to be unable to receive a further determiner: * 此马之千里者. The sequence 此马之千里者 does appear but it is almost always a nominal predication construction.

Box 2.24: What can be relativized, and possible external possession

若至 [力农畜，工虞商贾，为权利以成富，大者倾郡，中者倾县，下者倾乡里者]，不可胜数

It seems what is relativized here is the subject of the bracketed clause. But then what's the role of 大者倾郡? If we consider it to be a coordinated clause, then it seems an argument is moved from only one branch of a coordination construction: a clear violation of the coordinate structure constraint of extraction!

If we consider it to be a coordinated VP, then Classical Chinese should have a external possession construction: [商人]$_{\text{subject}}$ [大者倾郡]$_{\text{predicate}}$, in which 大者 is a part of 商人.

Or maybe this is a clausal pseudo-coordination: *what did Alex go to the store and buy.*

2.3.3 The determiner system

2.3.4 Prepositions

In Old Chinese, there are only two prepositions: 于 and 於. The exact usages of the two prepositions are not clear. In *Zuo Zhuan*, 于 is reserved for prepositional complements (§ 2.2.3.1.2), while 於 is for inter-predicate focalization (TODO). Other Old Chinese works have different conventions.

It is possible to omit the object of a preposition.

(82) 孔子 [因 —]$_{\text{reason}}$ 叹

2.4 Parts of speech division

Having had an overview of grammatical constructions in Classical Chinese, we turn to analyze the structure of its lexicon. That is to say, we now study the parts of speech division in Classical Chinese in the second sense in § 2.1.

Classical Chinese has no inflectional morphology for content words so it is not possible to define parts of speech based on inflections. Content words show much flexibility in their distributions in various syntactic environments, sometimes without any formal indications. These facts lead some to claim that Classical Chinese is a language without clear part of speech distinctions, so although we can talk about the

nominal or verbal usage of a root or a compound, strictly speaking we cannot talk about nouns or verbs, as there are no inherent lexical properties attached to roots that dictate their nominal or verbal usages. A more careful analysis, though, seems to reveal that at least some part of speech distinctions can be maintained in Classical Chinese, although Classical Chinese is much more tolerant to ad hoc re-categorization of roots than, say, English.

In principle, function words can be introduced together with their grammatical functions, but since the correct analyses of some constructions are still controversial and it may well be possible that the controversies reflect real historical linguistic divergence among speakers, function words are also discussed here for easier reference.

2.4.1 Nouns and verbs

A noun-verb distinction is supported by carefully examining traditionally called noun-used-as-verb phenomena (§ 6.1.1). If the lexicon of Classical Chinese contains *only* non-categorized roots, the interpretation of verbal usages of a word that usually appears in nominal environments should vary rather freely. What is actually attested however is not different from similar phenomena in other languages. In some cases, it seems a root is first categorized as a noun and then undergoes something similar to English *-ize* (albeit without any explicit marking), so only the nominal usage needs to be recorded as a lexical entry, but the lexicon controls whether a derivation step is viable. In other cases, both the nominal and verbal usages are to be recorded in the lexicon, as they cannot be inferred regularly from each other. In both cases, how a root is possibly categorized is stored in the lexicon, meaning that calling the nominal use of a root a *noun* and the verbal use of a root a *verb* is not problematic at all even in Classical Chinese. Sporadic ad hoc re-categorization of roots does exist, but this does not support the idea that part of speech division does not exist at all in the lexicon.

A terminological caveat is what appears as an argument is not necessarily a NP: it can be a complement clause. The main verb of a complement clause is not in a nominal position. Some may call complement clauses "nominal clauses", but this is misleading as the internal structure of a complement clause is not the same as that of a NP.

2.4.2 The adjective class

An adjective class can also be established in Classical Chinese, although its behavior is strongly verbal.

A caveat, similar to the caveat that an argument is not necessarily a NP, is that an attributive phrase is not always an adjective phrase. In existing modern studies, statements like "a verb used as an adjective" is usually avoided: wordings like "something is used as an attributive" are adopted instead.

Box 2.25: Traditional grammars

List some Classical Chinese grammars in which 动词作形容词 etc. never appear.

Box 2.26: A comprehensive list of Classical Chinese parts of speech

Noun, verb, adjective: any other content words?

2.4.3 Particles

Grammatical particles are not content words and in principle can be introduced together with the grammatical categories and relations they express. The long and complicated history evolution of Classical Chinese however means a particle may have multiple quite different uses possibly due to grammaticalization, so a surface form-to-function discussion on particles is of great descriptive value.

> **Box 2.27: Classification of particles**
>
> Do I need to classify particles?

者 The particle 者 most frequently appears as a relativizer, a complementizer, or in the *zhe-ye* construction. The three functions can be uniformly analyzed as the function of a low-level determiner (Aldridge 2009).

之 This

Chapter 3

Phonology and the writing system

Box 3.1: On the writing system

- 隶定和简化: one keeps the structure of a character and only alters the components, another messes up the structure
- 谐声
- Syntax within the character?

3.1 Theoretical consequences

Now we discuss the cognitive status of the "grammar of characters" sketched above. Questions relevant to this topic include whether grammar-like rules governing the structure of Chinese characters are synchronic or historical, and if they are synchronic, whether they derive from human's language capacity or from some other cognitive capacities. We note that the latter question is ultimately related to the big questions in theoretical linguistics and cognitive science (see also discussions at the end of § 1.3.1): if the grammar of Chinese characters mimics the grammar of spoken natural languages but the network in the brain processing Chinese characters is completely independent of the language network, then what we thought were unique to languages probably are not domain-specific to languages at all.

Regarding the first question, neurolinguistic experiments suggest that both holistic and sub-lexical processing exist in human brain. The existence of holistic processing is supported by the fact that Chinese readers find it easier to tell completely different characters apart than to tell characters with shared components apart, while the existence of sub-lexical processing is supported by the fact that characters with valid semantic or phonetic components are possessed more quickly (Duan and Cai 2024, § 2.2). So literate Chinese speakers do have the components (subconsciously) in mind when reading Chinese characters. Yet the same can be said for all orthographic systems (Myers 2019, pp. 23-25). English orthography, for instance, has phonology-like rules like *-y* + suffix → *-i*-suffix (Myers 2019, p. 26).

What makes Chinese characters special is that we have non-regular structures?

Regarding the neurolinguistic properties of Chinese characters, we note that the brain region primarily responsible for recognizing Chinese characters is the Visual Word Form Area, which is not a part of the language network in the brain, and this suffices to be an argument against the assumption that the "grammar" of Chinese characters and the grammars of spoken languages have the same neurological origin (Myers 2019, pp. 209-210). Still,

Chapter 4

Nouns

4.1 Derivation

The verbs 出 ('go out'), 入 ('enter'), 亡 ('die, decay') are regularly derived to 出 ('what goes out'), 入 ('what comes in') and 亡 ('what dies'). This derivation pattern however is not

> **Box 4.1: Deverbalization derivation**
>
> Summarize deverbal derivations.

4.2 Proper nouns

4.2.1 Personal names

Personal names in Classical Chinese have the surname-given name order. Yet the meanings of both *surname* and *given name* are not always clear. Before the Warring States period, the name of an aristocrat had two surnames: 姓 *xìng* 'ancestral clan name', 氏 *shì* 'branch lineage name'.

The merger between the two was eventually finished during the Han dynasty, possibly because of the collapse of the *fengjian* system: people who were not members of the Zhou dynasty nobility became politically active, who likely had neither ancestral clan names or branch linear names. It was common for them to adopt their birthplaces as their surnames, which did not fit into the double surname system, leading to the elimination of the distinction between *xìng* and *shì*. Sima Qian himself does not clearly distinguish between the two.

(1) 及生，名為政，姓趙氏

Besides the surnames, people also have 名 *míng* 'given name', the name given at birth. Still it is considered impolite to use *this* given name when having conversations with strangers or not-so-close friends: in the latter cases, another name, the 字 *zì* 'courtesy name (lit. character)' is used. A courtesy name, according to rites of Zhou, was to be given at the adult ceremony of a man, after which addressing him by his "regular" given name (i.e. his *míng*) was disrespectful.

(2) 幼名，冠字

A courtesy name may be formed by prefixation: it was common before the Qin dynasty to first have a monosyllabic proper of the courtesy name and then prefix 子 (a respectful title) or a numeric index (伯 'first', 仲 'second', '叔' 'third', '季' 'last') expressing the man's birth order in his family.

For example, an important disciple of Confucius, usually known as 子路 in the modern literature, has 由 as his name given at birth (i.e. his *míng*): 子路 is his courtesy name, where 子 is a respectful title for a man. In *Analects*, Confucius usually calls him 由 because their close relations, while others, including the editors of *Analects*, call him 子路 (i.e. his *zì*). His *shì* is 仲, and therefore when both the surname and the given name are called, 子路 is known as 仲由.

(3) 子路聞之喜。子曰：「由也好勇過我，無所取材。」

(4) 仲由可使從政也與

The practice to have courtesy names rapidly ceased in the twentieth century, when, quite similar to how the distinction between the ancestral clan name and the branch lineage name disappeared, due to rapid modernization, scholar-officials lost their positions in government, and rural land owners lost their properties both in mainland China and in Taiwan due to violent or non-violent land reforms. The "lay" citizens and peasants, now at the center stage of economy or even politics, had never haven serious exposure to the tradition of courtesy names, and it is predictable that no one cared about courtesy names anymore.

Moreover, besides the courtesy name, certain people may have art names, or 號 *hào* 'lit. mark'. A person may have multiple art names. The great poet 李白, whose courtesy name is 太白 'lit. extreme-white', had 青蓮居士 as his art name or *hào*.

4.2.2 Place names

TODO: e.g. 姑苏

Chapter 5

Pronouns

Box 5.1: Third person pronouns

之 seems to be the accusative pronoun in Old Chinese. 其 seems to be the genitive pronoun, and may be a phonological fusion of 之 and a possessive marker.

我，吾，

在下陛下阁下

Chapter 6

Verbal morphology

6.1 Stem derivation

6.1.1 "Nouns used as verbs"

The conventional term in Mandarin Chinese 名词作动词 'nouns used as verbs' covers two phenomena, corresponding to multiple functions and zero derivation (Dixon 2010, § 11.3), and also the rare case of ad hoc re-categorization of a root.

6.1.1.1 Multiple functions

Some roots have both nominal and verbal uses, and there is usually some semantic connection between the interpretations of the two uses, but this is not regularly inferrable. Here we consider some examples in Yang, Kong, and Zhou (1991):

- 楚 may mean 'the Chu state' or 'do what Chu people do'.
- 床 may mean 'bed' or 'settle down your bed or sleep on a bed'.
- 城 may mean 'city, castle' or 'build a city'.

The interpretation of the verbal usage is usually *not* decided from the meaning of the root and that the root is used in a verbal environment; rather, it is instructed by the lexicon. Therefore, the verbal usage of 城市 only means 'build a city' although the 'do city-related things' reading in principle could make sense.

Therefore, roots like 城, 楚 and 床 have double functions: nominal and verbal, but the two functions are likely not related to each other by regular grammatical rules. This corresponds to the "multiple function" case in Dixon (2010, § 11.3). Moreover, what is stored in the lexicon is not the bare, non-categorized root 城, but one noun lexeme 城 'city' that specifies its nominal usage and one verb lexeme 'build a city' that specifies its verbal usage, and other seemingly possible ways to categorize the root, although attested elsewhere, are ruled out by their absence in the lexicon.

The boundary between roots with double functions and roots undergoing zero derivation (see below) is somehow blurry, as the nominal and verbal uses of 城 and 床 still seem to show a common pattern and may be understood as a rare derivation. This blurriness leads many grammatical works on Classical Chinese to simply refer to the two phenomena uniformly as "nouns used as verbs".

6.1.1.2 Zero derivation

In other cases the meaning of the verbal use of a root usually appearing in a nominal context is regularly derived from the nominal meaning. This is because although tropative or causative

derivations in Classical Chinese are mainly verb-to-verb, they can also be applied to nouns. In this way from 臣 'servant, official, minister' we have the causative verbal usage 'make sb. dependent to', and from 客 'guest' we have the tropative usage 'consider sb. as a guest'. These verbal usages are nothing different from noun-to-verb derivation observed in other languages, so we regard the relevant phenomena as zero derivation as in Dixon (2010, § 11.3).

In zero derivation, the meaning of the nominal usage has to be recorded in the lexicon, the meaning of the verbal usage can be automatically decided from the derivation rule. These derivations are however not completely regular and not for every word: the lexicon also controls whether a derivational rule applies.

6.1.1.3 Ad hoc re-categorization

There are sporadic verbal usages of nouns that are almost never attested elsewhere, like 军 in 沛公军霸上. This means that ad hoc re-categorization of roots is possible in Classical Chinese, and the meaning is to be decided from the context. This is also possible in English (as in *I might [guinea pig] it for you.*) but usually not accepted in formal texts. Alleged ad hoc categorized Classical Chinese roots are indeed a possibility, after all, although their frequency is not high enough and cannot be exaggerated to be the norm rather than the exception.

6.1.2 Numeral-to-verb derivation

(1) 六王毕，四海一

6.1.3 Adjective-to-verb derivation

Box 6.1: Adjective-to-verb derivation

则其好游者不能穷也
山多石，少土
亲贤臣，远小人

6.2 Complex predication

Box 6.2: Complex predication

What matters: synchronic, or diachronic? Relation to valency alternation (no relation when diachronic)

Chapter 7

Verb valency

Box 7.1: More topics on argument structure

- Morphology?

7.1 Simple argument structures

7.1.1 Prototypical DO

The subject of a DO clause usually has to be animate, because it voluntarily initiates the event described by the clause (1). The subject is an *agent*, as opposed to a *causer* (§ 7.2) or a *theme* (§ 7.1.2).

(1) 桓公杀公子纠

7.1.1.1 Unique properties of DO verbs

Mei (2015, p. 272) lists some criteria to distinguish a transitive DO verb from a transitive CAUSE verb (§ 7.2.1).

We note that certain CAUSE verbs may gradually develop a lexicalized meaning and eventually get reanalyzed as a DO verb (Mei 2015, pp. 269-271).

7.1.2 Prototypical BECOME and BE verbs

7.1.2.1 The intransitive usage

A BE verb describes a state; a BECOME verb describes the change of a state. In both types of argument structures, the sole argument is a *theme*: the situation happening to it just happens, and usually it does not have much control over it nor any volition to trigger it (Li 2004, p. 345; Mei 2015, p. 275).

In Classical Chinese, just like in other languages, BECOME/BE verbs often have established causative usages, forming CAUSE-BECOME/BE argument structures with the *causer* argument being the subject and the *theme* argument being internal (§ 7.2.1). When a causer is absent, the structure of the clause is comparable to what sometimes is known as the middle voice in English (e.g. *the door opened*; cf. the transitive CAUSE-BECOME *I opened the door*). (2) is an instance: in its CAUSE-BE usage (10), the argument that is described as weak is an internal argument appears after the verb, but in (2), the argument that is described as weak is the *subject*: the internal theme argument gets promoted to the subject position.

(2) 秦强而赵弱

The "middle voice" construction exemplified in (2) (known as 内动 in Mei (2015)) has a subject, which corresponds to the argument that is the object in the CAUSE-BECOME/BE construction (i.e. the internal argument). It is however possible (although rare) for the subject position to be unfilled, and the internal argument remains in-situ. For instance, the verb 鸣 'chirp' appears in "middle voice" clauses (3), but its sole argument can also stay *after* the verb (4). The structure of (4) can only be reasonably conceived if we assume that 鸣 is a BE verb, denoting a state where bugs continue to make noise, and that the sole argument 蜩 remains in-situ and is not promoted to the subject position. No other analysis is available: for instance a DO verb can never have such a behavior (Mei 2015, p. 351).

(3) 蝼蛄 鸣
? chirp
'??? chirp.' (礼记・月令)

(4) [五 月]$_{\text{temporal}}$ [鸣]$_{\text{predicate}}$ [蜩]$_{\text{internal argument}}$
five month chirp cicada
'In the fifth (lunar) month, cicadas chirp.'

7.1.2.2 The alternation between BE and BECOME

Alternation between BE and BECOME verb frames is natural. Some BECOME verbs however do not have BE counterparts.

7.1.3 Non-conventional BE/BECOME clauses

Some verbs license subjects that look like arguments of prototypical BECOME or BE verbs: the subject may be animate but it does not volitionally trigger the event. The situation "just happens to be the case", and the subject can be described as a *theme* and not an *agent*. What sets them apart from prototypical BECOME or BE verbs in § 7.1.2 is the fact that the subject seems quite unlike an internal argument. In (5), the subject 火 'fire' is definitely a theme and not an agent: the fire does not get to *decide* if it burns the flag (Mei 2015, p. 276). Still the clause is not a prototypical BECOME one as there is an internal argument 其旗 in it, and the theme 火 is an *external* theme (Mei 2015, p. 353). These verbs therefore have difficulties participating in synthetic causativization (§ 7.2.1).

(5) 火焚其旗

7.1.4 Experience verbs

Some experience verbs, mostly verbs about emotions, behave like BECOME verbs (Mei 2015, p. 273): when used as transitive verbs, the subject do not look quite agentative and the clause is likely causative (7), and when used as intransitive verbs, there is a clear internal change-of-state meaning (6).

(6) 孔子成春秋，而乱臣贼子惧 (孟子・滕文公章句下)

(7) 惧之以怒 (左传・昭公十三年)

On the other hand, perception verbs (e.g. 见 'look') and cognition verbs (e.g. 知 'know') are often transitive, and therefore are not compatible with the synthetic causative construction (§ 7.2.1; Mei 2015, p. 274). Intuitively, these verbs are DO-like according to the criteria listed in § 7.1.1. For instance, they can appear in 所 construction (8).

(8) 異乎吾所聞

Certain perception verbs however have developed a figurative, fossilized meaning, and when intransitivized, can participate in synthetic causativization (§ 7.2.1, 11). This possibility indicates that these fossilized usages are BECOME- or BE-like: 见 'meet formally' therefore means 'in the state of regularly meeting an important figure'.

7.2 Various causative constructions

A *causer* makes a situation to be the case, but does not always do so intentionally. It can therefore be inanimate, as opposed to how an *agent* behaves (§ 7.1.1).

We can divide causative constructions in Classical Chinese into synthetic and analytic ones. In the synthetic causative construction, there is only one verb in the surface form: the causative valency alternation is supposedly marked by a prefix **s-*, which is invisible in the written texts but is reflected by tonal changes of the verb. If a root develops a lexicalized usage in the synthetic causative construction, then CAUSE verb is formed.

7.2.1 Synthetic causative

The synthetic causative construction applies to existing argument structures, or sometimes bare roots. The synthetic causative construction cannot be applied to a DO construction: the reason is probably because a DO construction is too "big", already having a full-fledged wannabe subject (Mei 2015, p. 363-364).

On the other hand, the syntactic causative construction can be applied to a pseudo-passive structure (9) and BECOME or BE (10) argument structures. Certain intransitivized experience verbs, possibly having an argument structure comparable to a BECOME/DO verb (§ 7.1.4) with a wannabe subject also have causative usages (11), but their transitive counterparts are never compatible with the synthetic causative construction (Mei 2015, p. 274).

Finally, the synthetic causative construction can be directly applied to a root (12): the word 妻 'wife' is sometimes used as a verb, meaning 'to marry daughter to ...', inconsistent with the meaning of (12). Therefore, in (12), 妻 is ad hoc categorized into a CAUSE verb, its usual verbal usage being irrelevant here.

(9) 是 夭 子蛮，杀 御叔……
this die.young NAME kill NAME
'This woman made Ziman die at a young age, and got Yushu killed...'

(10) 以 弱 天下 之 民
PURPOSE weak world GEN people
'...to weaken the people.'

(11) 子尾 见 疆
NAME formally.visit NAME
'Ziwei let Jiang formally visit (with Xuanzi).' (左传 • 昭公二年)

(12) 妻帝之二女

The labile S/O alternation between the BE/BECOME usage and the CAUSE-BE/BECOME usage is quite regular in Classical Chinese; verbs allowing this alternation are sometimes known as *ergative verbs* (Mei 2015, p. 378), although the phenomenon is about the core argument structure and has nothing to do with ergativity in alignment. It should be noted that not all BECOME/BE verbs are compatible with the synthetic causative construction. For instance, 鸣 'chirp' in (4) does not have a transitive CAUSE-BE usage. More examples are given in Mei (2015,

p. 276). On the other hand, some clauses that look like CAUSE-BE/BECOME clauses actually do not have BE or BECOME counterparts (§ 7.2.2).

7.2.2 Fossilization of synthetic causative construction

Some CAUSE verbs are fossilized, and do not have clear intransitive counterparts i.e. the un-causativized counterparts. For instance, 伤 'hurt' typically is a state transition verb meaning body, etc. being hurt, and it also has a causative (i.e. CAUSE-BECOME) meaning ('make ...hurt'). The CAUSE-BECOME verb frame of 伤 however has gained a separate lexicalized specific that can't be transparently inferred from the meaning of the BECOME usage: it can mean 'let ...be demaged', in which the object is not necessarily body or a person. This usage of 伤 has no BECOME or other intransitive counterpart. The absence of a BECOME counterpart can be proven by the ability for this fossilized figurative usage of 伤 to undergo "passivization" (13), which is otherwise not possible (§ 7.3.1).

(13) 女红伤则寒之原也

Verbs like 伤 in like (13) can easily be reanalyzed as DO verbs. This is likely a diachronic path of the creation of DO verbs. The verb 败 for example seems to be originally a BECOME verb ('to get corrupted') and have later gained a specific meaning of 'to defeat' in its CAUSE-BECOME usage, which had eventually evolved into a DO usage (Mei 2015, p. 285).

7.2.3 Analytic causative

Classical Chinese has an analytic construction to express the causative meaning. In (14), the word 使 is applied to the stative structure 渚者居中原 'people living on small lanbds reside inland', meaning 'let people living on small lands in water live inland'.

(14) 不 使 [渚 者]$_{\text{shared object: NP}}$ [居]$_{\text{BE}}$ [中-原]$_{\text{locative object: NP}}$
NEG let small.land.in.water NMLZ reside middle-land
'...do not let people living on small lands in water live inland'

The word 使 can be replaced by 令 (15) or 俾 (16) (Mei 2015, p. 376). Note that in (15), 令 seems to applied to a DO argument structure.

(15) 令军勿敢犯

(16) 俾民不迷

A question is whether 使 is a lexical verb, or just a non-incorporated causative marker. The fact that (15) involves a DO verb deviates from the behavior of the synthetic causative construction, which is not compatible with DO verbs (§ 7.2.1). Further, the fact that we can choose among 使, 令 and 俾 is rather unusual for a grammatical marker: we also note that 使 retains the meaning of 'send sb. to do sth.' and 令 retains the meaning of 'command sb. to do sth.' Therefore, we consider 使, 令 and 俾 to be *lexical verbs*, and not grammaticalized causative markers. What is grammaticalized is the causative clause with the form of V_1 NP VP, with V_1 being one of these verbs.

7.3 "Passive" constructions

7.3.1 The agent-less pseudo-passive construction

What is often known as the passive in Classical Chinese is not really a passive construction comparable to the English or Latin passive. The main problem is the lack of a grammaticalized

way to say the agent: in a "true" passive construction, the original subject is somehow demoted (represented by the appearance of *by* in English or the ablative case in Latin) and sometimes omitted, and an internal argument is promoted to the subject position, while the so-called "passive" constructions in Classical Chinese – which we name the *pseudo-passive* here – are *obligatorily removed* (Mei 2015, p. 287-289).

7.3.1.1 Passivization vs. information packaging

In a language that allows multiple topicalization and focalization, whether omission of the agent is to be analyzed as argument structure alternation instead of information packaging needs discussion. One may want to analyze the promotion of the patient argument to the subject position as pro-drop of the agent plus topicalization of the patient. This analysis essentially eliminates the necessity of the concept of subject in this construction. However, we contend that this analysis is not viable.

First, we note that Classical Chinese has explicitly marked passive constructions like the 见-construction. Their existence cannot be easily captured within an information structure-only framework.

Second, we have both passivization after causativization and causativization after passivization (§ 7.3.1.2), and the semantics and the word order in the two constructions are regularly determined. This strongly indicates that we are encountering valency alternation.

Third, information structure-driven omission of arguments, i.e. pro-drop in Classical Chinese mainly applies to arguments representing *known* entities. Yet the agent omitted in a pseudo-passive construction is frequently *unknown*.

7.3.1.2 Relation with causative constructions

It is not frequent but not impossible to apply pseudo-passivization to a causative construction (Mei 2015, p. 283). Suppose we have a bivalence causative construction. If pseudo-passivization is applied to it, we are to suppress the external argument and let the internal argument to be the subject. But such a bivalence causative construction usually has a CAUSE-BECOME structure, and removing the causer leaves us a clause that looks just like a BECOME clause. So there are two competing analyses, and since usually if a verb root is lexically licensed to head a CAUSE-BECOME clause, then its usage in a BECOME clause is also in the lexicon, the simpler BECOME analysis is preferred. However, where this preference is eliminated, pseudo-passivization of a causative construction is possible (Mei 2015, pp. 284,370-372; § 7.2.2).

Given the strict conditions of pseudo-passivization *after* causativization, alternate application of pseudo-passivization and causativization that results in a PASS-CAUSE-PASS-… argument structure is not attested, although such a combination is not semantically unfeasible (e.g. consider 'some one is made to be made to do sth.'), and is occasionally attested, probably in a playful way, in living languages like modern Japanese (17). This indicates that the argument structure in Classical Chinese is subject to considerable lexicalization and valency alternation constructions are not completely productive. In this particular case of the impossibility of a CAUS-PASS-CAUS chain, what is revealed probably also includes the decay of productive voice affixes in Old Chinese.

(17) 見つかると削除させられさせられる

mitsukar-u to sakujo-sase-rare-sase-rare-ru
see-PLAIN.NONPAST COND delete-CAUS-PASS-CAUS-PASS-PLAIN.NONPAST

'If it is discovered, someone will be made to be made to delete it.' (Modern Japanese, from Internet.)

7.3.1.3 The case with multiple internal arguments

When the pseudo-passive construction is applied to double object clauses, it is the *recipient* that is promoted to the subject position (18) (Mei 2015, p. 421). On the other hand, when the pseudo-passive construction is applied to the corresponding prepositional argument construction, it is the *theme* that is promoted to the subject position (19). These phenomena establish a hierarchy of *externality* of arguments (§ 2.2.5.1.2).

(18) 诸侯，赐弓矢然后征

(19) 药言先献于贵，然后闻于卑

TODO: 杀御叔

7.4 Experiential valency increasing

7.4.1 The affective constructions

Classical Chinese has two affective constructions, in which the subject is an experiencer suffering something bad from the situation described by the latter (Mei 2015, pp. 354-358).

The first affective construction simply attaches an experiencer to an argument structure (20). In (20a), 亡 appears as a BECOME verb (§ 7.1.2): it is intransitive and its subject, the *Odes*, did not have control over its being ignored. In (20b), a new argument – the experiencer subject – is introduced to the argument structure of 亡: the meaning of the sentence is 'the shepherds suffered from the sheeps getting lost.'

(20) a. 亡 as a BECOME verb

[诗]$_{\text{subject,theme: NP}}$ [亡]$_{\text{BECOME}}$ 然后 春秋 作
poem **get.lost** then Spring-Autumn compose

'The *Odes* got lost, and then the *Spring and Autumn Annals* was composed.'

b. 亡 as a AFFECTIVE-BECOME verb

[二 人]$_{\text{subject: NP}}$ [相 与 牧 羊， 而 俱 [亡]$_{\text{AFFECTIVE-BECOME}}$
two person mutually go.together herd sheep CONJ all **get.lost**
其 羊]$_{\text{coordinated VP}}$
POSS sheep

'Two people herded their sheep together, and they both lost their sheep (lit. suffer from their sheep's missing).'

The second affective construction *obligatorily* has an object, which is in possession of the subject. The meaning of (21), for example, is that Confucius was frustrated by the fact that his tree was cut in Song.

The fact that the object (树 and 胁 here) should not contain any possessive markers and has to be interpreted as something being possessed by the subject suggests that the second affective construction is an external possession construction.

(21) 吾再逐於鲁，伐树於宋

(22) 范睢折胁於魏

Box 7.2: External possession

External possession as subject

7.4.2 Tropative

Tropative is a construction which attaches an experiencer to a BE argument structure, with the meaning being '*A* consider *B* to be ...' (Mei 2015, pp. 413-414) The Classical Chinese tropative is actually not limited to stative verbs: it also applies to nouns. It is however not likely that this construction comes from transformation of the nominal predicate construction (§ 2.2.1). The main difference is that in the nominal predicate construction, the predicate is a noun *phrase*, but the tropative construction never takes a nominal predicate as input.

7.5 Applicative constructions

Box 7.3: Applicative constructions

benefactive; the claim that double object constructions are similar to benefactive constructions; what object gets passivized. See Mei (2015, p. 421).

Chapter 8

Tense, aspect, modality

8.1 Time adverbs

It is probably more appropriate to recognize these TAM markers as (semi-)lexical adverbs instead of auxiliaries or particles, because the class of TAM markers seems to be open during the Classical period. We believe that they have been incorporated into a grammaticalized TAM system, because depending on their meanings, they have a fixed linear order, consistently with the order of TAM markers observed cross-linguistically (§ 2.2.4). For instance, from (1), we observe that the epistemic adverb 或 appears before the future adverb 將. Still, multiple time adverbs are rare in truly Classical texts, and whether an adverb is a part of the TAM system or a temporal peripheral argument like *at that time* sometimes has to be determined by its semantics.

Box 8.1: List of time adverbs

list

- Are some of them peripheral arguments?

8.1.1 Subject-oriented adverbs

The pre-verbal particle 或 seems to be an epistemic modality marker. In (1), we see the coexistence of the epistemic 或 and a future adverb 將.

(1) 或將豐之，不亦難乎

必 indicates the speaker's strong confidence towards the statement being made; similar adverbs include 固, 實, and 誠 (Li 2004, pp. 376-377). These adverbs also appear before the future 將, further confirming their status as epistemic modality or evidentiality markers.

(2) 知惠之必將至也

(3) 諾，吾固將圖之

We can further observe both the linear order 固 > 必 and the reverse.

(4) 固必通乎性命之情者

(5) 將欲弱之，必固強之

8.1.2 The tense adverbs

嘗 seems to be a marker of the simple past.

The concept of future events can be expressed by 將, but it is not compatible with 嘗. On the other hand, in English, we have *would do sth.* or even *would have done sth.* The incompatibility of 嘗 and 將 is evidence for the syntactic distinction between how future is expressed in English and in Classical Chinese: in English, we have both a past/non-past distinction and a future/non-future distinction, beyond the standard Reichenbach system (Vikner 1985), while in Classical Chinese, the past future tense is non-existent and we have a tripartite past/present (related to 矣; § 8.3.1)/future distinction.

8.1.3 The modality adverbs

(6) 又將能忍吾子乎

8.2 (Semi-)auxiliaries

In English, it is possible to have both *want to be able to do sth.* and *be able to want to do sth.*[1] A corpus search reveals that 可 'should, ought to' always appears before clearly lexical verbs, while the reverse order is not attested.

(7) 見其可欲也，則必前後慮其可惡也者

Further, consider 罪當可赦: 當 > 可.

8.3 The role of sentence final particles

8.3.1 矣 and the tense system

The appearance of the sentence final particle 矣 seems to have a tense effect. When 矣 appears and there are no other TAM markers, the relation between the event time and the speech time is still largely arbitrary, but the event always seems to have some relevance to the speech time: the meaning of English *somebody did something* or *somebody had did something* is not admissible. It seems that these observations can be summarized into the follows: the grammar of Classical Chinese does (although implicitly) recognize Reichenbach's speech time, reference time and action time, and when 矣 appears, the latter two are identical, hence prohibiting the simple past meaning (Mei 2015, pp. 437-446). However, after considering its combinations with other markers, we do not consider 矣 itself to be a tense marker (§ 10.2).

[1] This may sound awkward, but it makes sense in certain contexts, like *you have to be a special kind of nerd to be able to want to drive 28 hours for a video game.*

Chapter 9

Negation

Chapter 10

Sentence final particles

10.1 Sentence final particles as discourse markers

At the first glance, the sentence final particle 矣 marks the perfect aspect (1), while 也 is for clauses describing something happening regularly, not a single concrete event (2) (Mei 2015, pp. 443-445).

(1) 余助苗长矣

(2) 將发命也

矣 and 也, however, are different from prototypical aspect markers in several aspects. First, 矣 appears predominantly in direct quotations in Classical texts, which suggests that it has conversational functions. 也 frequently appears in narratives as a part of the judgemental construction (§ 2.2.1.2), which lacks TAM marking, and 也 cannot be a prototypical aspect marker in that context. Second, it seems that 矣 and 也 can be shared by two coordinated conjuncts with different subject (§ 2.2.2.4), which is rather unusual for an aspect marker.

Therefore, the grammatical category corresponding to 矣 and 也, whatever it is, is not a typical TAM category, and hence we disagree with Mei (2015)'s analysis. Its scope is wider than TAM categories: a TAM category is in relation with a nucleus clause, while 矣 and 也 are in relation with a sentence, i.e. an arbitrarily complex clause that is one utterance in a conversational context. This is consistent with the usual analysis of sentence final particles in modern Standard Mandarin (Paul 2014b; Pan 2021).

(3) 亦各言其志也已矣

Box 10.1: Alternative analysis

Zhu (2009, p. 233) acknowledges the wide spread of the analysis that sentence final particles are in relation with the whole sentence, not the nucleus clause, but insists that certain sentence final particles are a part of the predicate. Yet no convincing argument is provided. Among the three distributional classes he recognizes in Mandarin, two (marking the interrogative/imperative force, and attitude of the speaker) are uncontroversially attached to *sentences* and not nucleus clauses. The remaining class, which is called the "tense" class by Zhu (2009) and is structurally the innermost, resembles the class of 矣 and 也 discussed here, seems to be forbidden in most embedded clauses (Deng 2010), just like the other two class do. Therefore all the three classes of Mandarin sentence final particles described in Zhu (2009) are indeed in relation with the sentence and not the nucleus clause, which is consistent with the structural status of sentence final particles in Classical Chinese.

We also note that it is not completely impossible for two independent nucleus clauses

to share one TAM marker. In Japhug, for example, a series of nucleus clauses with different subjects can be coordinated with the TAM categories being marked at the end of the compound clause (Jacques 2021, pp. 1090-1091). However, Japhug lacks clear clause-level subject (Jacques 2021, § 2.5.3; although subjecthood is well-defined at the level of argument structure (Box 2.12)), and therefore coordinated nucleus clauses with a shared TAM marker but different "subjects" is less strange in Japhug than it is in Classical Chinese: in the latter, we have a well-defined clausal pivot grammatical relation (§ 2.2.5.2) whose scope is over all TAM categories, making clauses sharing the TAM marker but not subjects rather unusual, but in the former this is probably not the case.

Mei (2015, pp. 443-445) relies solely on semantic criteria. Although we do not believe his analysis of 也 and 矣 as tense marker is completely correct, the two clearly have non-trivial interaction with TAM categories, a phenomenon also observed in Mandarin, where the lowest sentence final particle has access to TAM categories of the nucleus clause (Paul 2014a, p. 258).

10.2 矣

矣 is not a tense marker Besides the fact that sentence final particles are discourse-oriented and therefore not prototypical TAM markers, we also note the following two facts.

First, 矣 is compatible with both 將 and 嘗 (4, 5; § 8.1.2), which probably means it is not directly related to the tense system.

(4) 難將至矣

(5) 牛山之木嘗美矣

Second, 矣 seems to have a weak exclamative function. For example, in (6), the speaker has something to say but has waited for quite a while for a change to say it. In (7), Confucius is impressed by how well cultivated 韶乐 '*Shao* music' is.

(6) 吾欲言之久矣

(7) 尽美矣

We therefore conclude that 矣 is not a tense marker, and instead analyze it as a "sentential aspect" marker, which informs the listener that a new piece of information arrives. This is consistent with the function of the sentence final 了 in modern Standard Chinese.

Box 10.2: Alternative analysis

In Ch. 11 of Mei (2015), it is claimed that 矣 is a prototypical tense marker. We agree with Mei (2015) in that Classical Chinese has a (rather weak) tense system, and we basically reiterate his arguments above. His arguments for the claim that 矣 itself is a tense marker however we do not consider valid. On p. 447, he argues that 矣 appears in imperative sentences and therefore cannot be related to aspectual categories. But the category of tense also rarely appears in imperative sentences. On the other hand, the sentential aspect category is essentially a discourse-level category, and has no inherent conflicts with the imperative speech act. On p. 448, he notes that in coordination, when 矣 appears at the end of the first clause, it sets up a reference time for the following clauses. Yet he states clearly that the reference time is actually the event time of the first clause, whose relation with the speech time is *not* specified by 矣.

On the other hand, if we accept the analysis that 矣 hints at new information that the speaker wants to express to the listener at the speech time, then we can easily un-

derstand why 矣, because a generic fact like *the Earth goes around the Sun* rarely is new information, and a piece of new information usually involves a specific situation at a specific time, which is inevitably compared with the speech time, and hence the tense system. The fact that 矣 is often related to the present tense or the present perfect tense may simply be because of the absence of other TAM markers: cf. (4, 5).

Other sentence final particles, like 也, do not have this effect: hence an alternation between narrations about events and descriptions on what these events mean on the battlefield (Mei 2015, pp. 444-445).

Box 10.3: The relation between topic and sentential aspect

A question is whether we can claim that the category of the sentential aspect is somehow higher than the category of topic. If so, then the sentential aspect belongs to CP. See Mandarin peripheral construals at the syntax-discourse interface, page 30. 是北京昨天下雪了：the focus can be on 北京 and 昨天下雪了 modifies 北京, or 了 modifies

Mei (2015, p. 445) and Pulleyblank (1995, p. 19) both claim that form 也矣 is impossible. It turns out that it is rare, but not non-existent. It however has to be admitted that 也已矣 is much more frequent than 也矣. Pulleyblank (1995, p. 20) analyzes 已 (when used as a sentence final particle) as a fusion of 也 and 矣.

(8) 是善惡之分也矣

Box 10.4: Other combinations

寡人之於國也，盡心焉耳矣

Chapter 11

Genres and formulae

11.1 Idioms

(1) 非请勿入

11.2 Classical-like formal documents

11.2.1 Background

Box 11.1: Background

- 各代公文
- 晚清民国

11.2.2 Pseudo-Classical documents

The text shown in (2) is an article from the Chinese version of Geneva Conventions, an example of the pseudo-Classical Chinese used in official documents. The text is Classical-like when we inspect the grammatical markers appearing in it: 之 in place of 的 as a relative clause marker marker and also a possessive marker, 是 in place of 这 as a demonstrative, and 彼等 in place of 他们 as a third person pronoun. Yet several features can be immediately noticed that deviate from the Classical standard.

(2) 战俘不得放弃本公约或上条所述之特别协定——如其订有是项协定——所赋予彼等权利之一部或全部

First, polysyllabic content words are prevalent. The verb 放弃 'to give up' is not unheard of in Classical texts (cf. 3), but it is usually analyzable as a verb-level coordination of 放 'to put down' and 弃 'to abandon'. 放弃 as a whole with the meaning of giving up something abstract is a feature of Modern Standard Mandarin, where 弃 is no longer acceptable when used alone as a verb, which is indeed the case in the Chinese version of Geneva Conventions.

Nouns like 战俘 or 协定 are never attested in common Classical texts. Their structures can be explained with the Classical grammar. Still, the fact that roots within them do not seem to appear freely suggests that the translator likely had the lexicon of Mandarin in their mind.

(3) 放棄詩書，極意聲色

Second,
It is not impossible for Classical grammar to appear in a pseudo-Classical document, though.

(4) 有组织之抵抗运动人员之在其本国领土内外活动者

11.2.3 Formulae

Box 11.2: Formulae

- 此令

11.3 Traditional letters

Box 11.3: Formulae

- 钧鉴
- 谨呈
- 母亲大人膝下，敬禀者，日前寄上海婴照片一张，想已收到

  The last example can be parsed with Classical grammar: ‘ At the knees of my mother, what is respectively reported below is: …’ Yet the formula X 鉴 is kind of hard to parse. Further, consider 道席 in 某公道席: it is similar to an epithet, or something like *your Majesty*.

  So we have three types of 提称语: one is related to the request that the recipient reads the letter (the X 鉴 formula), one is related to the profession or status of the recipient (like 道席), and the other is essentially a locative expression (膝下). The origin of these formulae remains unknown.

Chapter 12

Discussions on quirky examples

(1) [良人者]$_{\text{subject}}$ [所仰望而终身]$_{\text{predicate}}$ 也

It seems subjects of ordinary verbal clauses cannot be topicalized (2), but if the VP is emphasized, topicalization is possible.

(2) a. 三王既以定法度
b. * 三王，既以定法度

(3) 秦，虎狼之国，不可信，不如毋行

What's the structure of the sentence below?

(4) 將者，欲伐而未成，见其臣尚可以谏，而季氏尚可以止也

In the example below, the subject of the second clause is 'other people', and yet it seems the three clauses share a topic (i.e. 人).

(5) 人而不仁，疾之已甚，亂也

What's the structure of the sentence below? 'But as for Qiu, is it about state affairs?'

(6) 唯求则非邦也與

The sentence below seems to be due to omission: 宗庙会同，非诸侯而何为之？

(7) [宗庙会同]$_i$，非诸侯而何 $-_{\text{main verb}}$ $-_i$

(8) 赤也为之小，孰能为之大

(9) 其为人也孝弟，而好犯上者

(10) 且夫水之积也不厚，则其负大舟也无力

(11) (人) 如礼何

(12) 如之奈何

(13) 父母唯其疾之忧

References

Aldridge, Edith. “Generative approaches to ergativity”. In: *Language and Linguistics Compass* 2.5 (2008), pp. 966–995.

Aldridge, Edith. “Old Chinese determiner zhe”. In: *Historical syntax and linguistic theory* 18 (2009), pp. 233–250.

Baxter, William H. *Old Chinese: A new reconstruction*. Oxford University Press, 2014.

Boston, Marisa Ferrara, Hale, John T, and Kuhlmann, Marco. “Dependency structures derived from minimalist grammars”. In: *The mathematics of language*. Springer, 2009, pp. 1–12.

Bruening, Benjamin. “The lexicalist hypothesis: Both wrong and superfluous”. In: *Language* 94.1 (2018), pp. 1–42.

Cinque, Guglielmo. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press, 1999.

Danckaert, Lieven. “On the left periphery of Latin embedded clauses”. PhD thesis. Ghent University, 2011.

DeLancey, Scott. “The origins of Sinitic”. In: *Increased empiricism: Recent advances in Chinese linguistics*. John Benjamins Publishing Company, 2013, pp. 73–100.

Deng, Siying 邓思颖. “汉语句类和语气的句法分析”. In: 汉语学报 1 (2010), pp. 59–63.

Dixon, Robert MW. *Basic linguistic theory volume 1: Methodology*. Vol. 1. OUP Oxford, 2009.

Dixon, Robert MW. *Basic linguistic theory volume 2: Grammatical topics*. Vol. 2. Oxford University Press on Demand, 2010.

Duan, Xufeng and Cai, Zhenguang G. “Chinese Character Processing”. In: *Oxford Research Encyclopedia of Linguistics*. 2024.

Grimm, Nadine. *A grammar of Gyeli*. Language science press, 2021.

Harbsmeier, Christoph. “Irrefutable Conjectures. A Review of William H. Baxter and Laurent Sagart, Old Chinese. A New Reconstruction”. In: *Monumenta Serica* 64.2 (2016), pp. 445–504.

He, Leshi 何乐士. *Shǐjì yǔfǎ tèdiǎn yánjiū* 史记语法特点研究 *[A study on grammatical features of Historical Records]*. Shāngwù Yìnshūguǎn, 2005.

Hill, Nathan W. *The historical phonology of Tibetan, Burmese, and Chinese*. Cambridge University Press, 2019.

Huddleston, Rodney and Pullum, Geoffrey K. *The Cambridge Grammar of the English Language*. Cambridge University Press, 2002.

Jacques, Guillaume. *A grammar of Japhug*. Vol. 1. Language Science Press, 2021.

Jacques, Guillaume. "A possible trace of verb agreement in Tibetan". In: *Himalayan Linguistics* 9.1 (2010).

Jacques, Guillaume. "Agreement morphology: the case of Rgyalrongic and Kiranti". In: *Language and Linguistics* 13.1 (2012), pp. 83–116.

Jacques, Guillaume. "Tangut, Gyalrongic, Kiranti and the nature of person indexation in Sino-Tibetan/Trans-Himalayan". In: *Linguistics Vanguard* 2.1 (2016), p. 20150033.

Jacques, Guillaume. "The origin of the causative prefix in Rgyalrong languages and its implication for proto-Sino-Tibetan reconstruction". In: *Folia Linguistica* 49.Historica-vol-36 (2015), pp. 165–198.

Keidan, Artemij. "Subjecthood in Pāṇini's grammatical tradition". In: *Studi in onore di Giuliano Boccali* (2017), p. 107.

Kiparsky, Paul. "On the architecture of Pāṇini's grammar". In: *Sanskrit Computational Linguistics: First and Second International Symposia Rocquencourt, France, October 29-31, 2007 Providence, RI, USA, May 15-17, 2008 Revised Selected and Invited Papers*. Springer. 2009, pp. 33–94.

Li, Zuofeng 李佐丰. *古代汉语语法学*. 商务印书馆, 2004.

List, Johann-Mattis et al. "Vowel purity and rhyme evidence in Old Chinese reconstruction". In: *Lingua Sinica* 3 (2017), pp. 1–17.

Matisoff, James A. "Notes on Fang-Kuei Li's 'Languages and dialects of China'". In: *Journal of Chinese Linguistics* 1.3 (1973), pp. 471–474.

Matthews, Peter Hugoe. *What Graeco-Roman grammar was about*. Oxford University Press, 2019.

McCraw, David R. "An 'ABC' Exercise in Old Sinitic Lexical Statistics". In: *Sino-Platonic Papers* 202 (2010), pp. 1–42.

Mei, Guang 梅广. *Shànggǔ hànyǔ yǔfǎ gāngyào 上古漢語語法綱要 [A sketch of Old Chinese grammar]*. Sānmín Shūjú 三民書局, 2015.

Myers, James. *The grammar of Chinese characters: Productive knowledge of formal patterns in an orthographic system*. Routledge, 2019.

Ostapirat, Weera. "Notes on phonological correspondences and vocabulary distribution". In: *The peopling of East Asia: Putting together archaeology, linguistics and genetics*. RoutledgeCurzon New York, 2005, pp. 107–131.

Oxford, Will. "Inverse marking as impoverishment". In: *Proceedings of WCCFL*. Vol. 34. 2017, pp. 413–422.

Pan, Victor Junnan. "Sentence-final particles in Chinese". In: *Oxford research encyclopedia of linguistics. Oxford: Oxford University Press. Published online on* 28 (2021).

Paul, Waltraud. *New perspectives on Chinese syntax*. Vol. 271. Walter de Gruyter GmbH & Co KG, 2014.

Paul, Waltraud. "Why particles are not particular: Sentence-final particles in Chinese as heads of a split CP". In: *Studia linguistica* 68.1 (2014), pp. 77–115.

Pulleyblank, Edwin George. *Outline of classical Chinese grammar*. Ubc Press, 1995.

Sagart, Laurent. “A model of the origin of Kra-Dai tones”. In: *Cahiers de Linguistique Asie Orientale* 48.1 (2019), pp. 1–29.

Sagart, Laurent. “The higher phylogeny of Austronesian and the position of Tai-Kadai”. In: *Oceanic Linguistics* 43.2 (2004), pp. 411–444.

Sagart, Laurent. “The origin of Chinese tones”. In: *Proceedings of the Symposium/Cross-Linguistic Studies of Tonal Phenomena/Tonogenesis, Typology and Related Topics.* Institute for the Study of Languages, Cultures of Asia, and Africa, Tokyo ⋯ 1999, pp. 91–104.

Sagart, Laurent. “The roots of old Chinese”. In: (1999).

Sagart, Laurent et al. “Dated language phylogenies shed light on the ancestry of Sino-Tibetan”. In: *Proceedings of the National Academy of Sciences* 116.21 (2019), pp. 10317–10322.

Shuya, Zhang, Jacques, Guillaume, and Yunfan, Lai. “A study of cognates between Gyalrong languages and Old Chinese”. In: *Journal of Language Relationship* 17.1-2 (2019), pp. 73–92.

Siddiqi, Daniel. *Syntax within the word: Economy, allomorphy, and argument selection in Distributed Morphology*. Vol. 138. John Benjamins Publishing, 2009.

Sidwell, Paul and Jenny, Mathias. *The Languages and Linguistics of Mainland Southeast Asia.* De Gruyter Mouton, 2021.

Vikner, Sten. “Reichenbach revisited: one, two, or three temporal relations?” In: *Acta linguistica hafniensia* 19.2 (1985), pp. 81–98.

Yang, Zhaowei 杨昭蔚, Kong, Lingda 孔令达, and Zhou, Guoguang 周国光. *Gǔhànyǔ cílèi huóyòng cídiǎn 古汉语词类活用词典 (A dictionary of word class changes of Classical Chinese).* 三环出版社, 1991.

Zhang, Menghan et al. “Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic”. In: *Nature* 569.7754 (2019), pp. 112–115.

Zhengzhang, Shangfang. “Decipherment of Yue-Ren-Ge (song of the Yue boatman)”. In: *Cahiers de linguistique-Asie orientale* 20.2 (1991), pp. 159–168.

Zhu, Dexi 朱德熙. “语法讲义”. In: 朱德熙文集， 第一卷. 商务印书馆, 2009.