

Mandarin Chinese notes

Jinyuan Wu

May 11, 2025

Chapter 1

Grammatical overview

1.1 Morphological typology

Mandarin lacks prototypical inflectional morphology but has rich derivational morphology. Compounding is the most frequent morphological device, and partly due to influences of European languages, partly due to grammaticalization, affixation is also frequently seen. Plus, reduplication plays an important role in Mandarin verbal and adjectival derivation.

One caveat about talking about derivational morphology is that it assumes the existence of a well-defined wordhood. It's often claimed that Mandarin lacks the word/phrase distinction. In this note, we will show that wordhood can indeed be clearly defined by syntactic, morphological and phonological standards. What makes Mandarin unique is that the three definitions of wordhood do not always overlap.

Box 1.1: Different standards of wordhood

Define

Because of this, Chinese lexicography is usually based on *morphemes*,

1.2 Clauses

1.2.1 Clauses are made of nucleus clauses and high-level categories

A clause can be divided into several clauses linked by **clause linking** constructions,¹ including **coordination** and **subordination**. Mandarin has ample information mark-

¹Huddleston and Pullum (2002) uses the term *sentence* to refer to a natural unit in dialogue, which I refer to as a *utterance*. The term *sentence* here refers to a clause that qualifies as an utterance.

Some people, like Deng (2010, p. 140) as well as Dixon (2009), use the term *clause* for subject-predicate constructions with no speech force marking. (Deng (2010) uses 句子 as the Mandarin counterpart of *sentence* and 小句 as the counterpart of *clause*.) In this way, sentence-final particles strictly shouldn't be regarded as a part of the clause, and they may be discussed together with other higher level constructions like clause linking. This notion of clause correctly highlights the hierarchy in clausal structures. The problem with this terminology however is that in traditional grammars, the term *clause* does refer to units that have SFPS.

This note therefore refers to all units larger than the subject-predicate construction as clauses, which may or may not be sentence. The subject-predicate construction is instead named the *nucleus* clause. The internal complexity of a clause is still relevant for example in clause combining.

ing phenomena, and thus a clause can be divided into one or more **topics**, if any, and a **comment**, the latter being the **nucleus clause** plus possible **sentence final particles** (1). Note that topicalization and clause linking can happen successively (2), and coordination can also happen inside the nucleus clause (TODO: ref).

- (1) [张三]_{topic,i}, [[他]_{subject,i} 就 是 [个 王八蛋]_{copular complement}]_{nucleus clause}
 NAME 3 just be CLS turtle-egg
 罢了!
 SFP
 ‘Zhang San is a son of a bitch!’
- (2) [我]_{topic,i} [幸亏 —_i 昨天 没 来]_{nucleus clause}, [否则 —_i 就
 2 fortunately yesterday NEG come otherwise then
 被 困 住了]_{nucleus clause}
 BEI trap V2 ASP
 ‘Fortunately, I didn’t come yesterday, or otherwise I would have been trapped.’

1.2.2 Subject and predicate

The nucleus clause contains a subject (if any) and what is often known as a predicate,² which usually is a (extended) verb phrase but may also be a nominal. In the Mandarin simple nucleus clause, the definition of the subject, as opposed to the topic, is not trivially clear. Here we note that the nucleus clause has a neutral structure (§ 1.2.2.1), in which a subject appearing at the initial is both the argument structure pivot (§ 1.2.2.2) and the clause-level pivot.

1.2.2.1 The neutral order

The notion that in Mandarin, subject is the same as topic is prevalent. Taking one step further, one may argue that Mandarin has no argument structure at all and the word order in a clause is shaped by only information structure (LaPolla 2009). This grammar rejects this analysis.

First, we note that a information structure neutral order can be defined for most, if not all, clauses. An example is provided in (3). The two arguments, 饭 and 吃, can be reordered in a seemingly free way depending on their topicality, violating the common generalization that Mandarin has a SVO order. We however note that (3d) is completely unacceptable with the intended meaning. Playing with more possible orders, and we will find that the arguments seem to be only permitted to move *leftwards* (and thus 3d is not possible), consistent with the assumption that a neutral ordered nucleus clause is formed first, followed by topicalization. By analyzing subtle pragmatics differences, we find (3a) seems to be the “neutral” order (although it imposes weak topicality to 你 ‘you’, and 吃饭 ‘eat (lit. eat meal)’ is focalized).

- (3) a. 你 吃 饭 了 吗
 2 eat meal SFP SFP

²Dixon (2009) argues against the definition of *predicate* as the main verb (or adjective) plus somehow “internal” arguments. He uses the term *predicate* to refer to the verbal complex instead. However, since I will need to compare the topic-comment construction with the inner structure of the nucleus clause, the term *predicate* will still be used in the way Dixon (2009) dislikes, because it’s the counterpart of the comment role in the topic-comment construction.

- ‘Have you eaten?’
- b. 饭 你 吃 了 吗
meal 2 eat SFP SFP
‘Have you eaten?’
- c. 你 饭 吃 了 吗
2 meal eat SFP SFP
‘Have you eaten?’
- d. *饭 吃 你 了 吗
meal eat 2 SFP SFP
‘Intended meaning: have you eaten? (Actual meaning: has meal eaten you?)’

We also note that there is no dangling topic in Mandarin (§ 1.5). This means *all* topics originate from somewhere within the nucleus clause. On the other hand, the subject, if well-defined by the usual pivot tests, is a part of the nucleus clause, and therefore in Mandarin, topic and subject are different.

1.2.2.2 Subject as pivot of argument structure

Being the initial constituent³ in clauses like (3a) has a clear relation to being the most prominent or the most *external* argument – the agent or causer or the patient in passive constructions. Examples like (4) can be explained by valency alternation.

- (4) 茶泡好了

1.2.3 The predicate

(5) is an illustration of a complicated nucleus clause. Its constituent structure is shown in Fig. 1.1, following the notation in Huddleston and Pullum (2002). We need to warn that the main information contained in Fig. 1.1 is the *scopes* of constituents surrounding the core verb phrase, while the function labels (e.g. *head*) and the form labels (e.g. *extended VP*) in Fig. 1.1 may be misleading, as 能在我的办公室跟你讨论一下 and 可能能在我的办公室跟你讨论一下 are both labeled as extended VPs, but clearly they have slightly different syntactic statuses: the auxiliary 可能 can be attached to the former but it can never appear twice and hence cannot be attached to the latter.

- (5) 我 [明天 可能 能 在我的 办公室 跟 你 [讨论
1 tomorrow AUX:possible AUX:ability at my POSS office with 2 discuss
一下]coreVP]extended VP
a.little.bit

‘Tomorrow possiblity I can have a discussion with you in my office.’

³Note that it is possible that certain constituents, like temporal constituents, naturally appear before the subject.

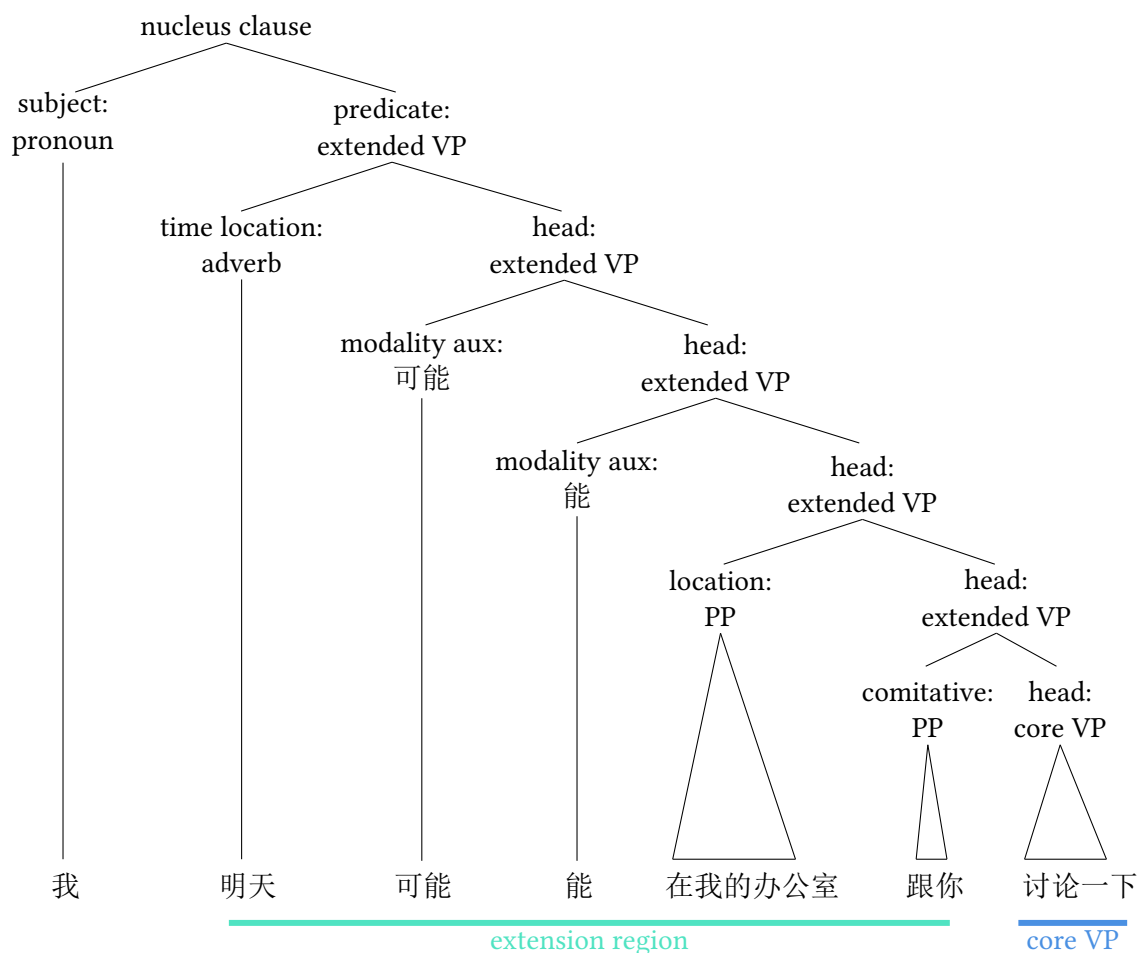


Figure 1.1: Tree diagram of (5)

It can be seen that when we have verbal prediction, the full, extended VP following the subject can be divided into an extension region and the core VP (§ 1.2.4). The extension region contains TAME auxiliaries and adverbs not realized in the verbal complex (§ 1.2.3.1), and peripheral arguments like temporal and spatial locations (§ 1.2.3.2). Sometimes the object may be fronted and it's also possible that a prepositional complement is fronted to this region.

It should be noted that in the disposal and passive constructions, the manner phrase may appear *after* the auxiliary (6), and in this case the boundary of the core VP can't be clearly defined at the surface level, which shouldn't be surprising as we do not expect to always see a clear-cut argument/adjunct distinction.

- (6) 我 明天 可能 能 在 我 的 办 公 室 跟 你 [把]_{auxiliary}
 1 tomorrow AUX:possible AUX:ability at my POSS office with 2 BA
 这 个 问 题 [好好]_{manner} 讨 论 一 下
 this CLS problem good discuss a.little.bit
 'Tomorrow possibility I can have a good discussion of this problem with you in my office.'

1.2.3.1 Tense, aspect and modality marking

In (5), it can be clearly seen that Mandarin has modal auxiliaries: the order of 可能 and 能 is strictly 可能 > 能 and never the inverse, suggesting that these modality markers

are grammaticalized items. More analytic markers of TAME categories can be found: it seems 据说 is a periphrastic marker of evidentiality, for instance: in (7), the order of the TAME markers is always 据说 > 可能, and not the inverse, suggesting that 据说 is a part of the TAME grammatical hierarchy.⁴

- (7) [这 辆 车]_{subject:NP} [据说]_{evidentiality} [可能]_{modality} 不 太 靠谱
 this CLS car is.said AUX NEG very reliable
 ‘It is said that this car may not be very reliable.’

Mandarin also has (semi-)inflectional marking of aspect by 了, 着 and 过 in the verbal complex (§ 1.2.4).

Whether Mandarin has something comparable to tense in more prototypical tensed languages is not clear. An observation is that Mandarin speakers often do not fully sub-consciously acquire the tense category when learning tensed languages like English. This, however, does not fully exclude the possibility of an impoverished tense system. TODO: ref

1.2.3.2 Peripheral arguments

The term *peripheral argument* is from Dixon (2009). We intentionally use the term here instead of the more frequent *adjunct*, because there are both TAME adjuncts and circumstantial adjuncts, the latter known as peripheral arguments in Dixon (2009).

A clear distinction between core and peripheral arguments, more often known as the argument/adjunct distinction, is not always possible. Some criteria used for the distinction are about structural closeness of the argument to the main verb, or in other words scope: the manner expression usually has scope over the core verb phrase, and thus the former is classified as a peripheral argument. Other criteria are based on licensing: intransitive use of a transitive verb is prohibited by the lexicon, or, in more technical terms, the verb root appearing in a verbal environment but without transitivity is not allowed by the lexicon (Siddiqi 2009). Thus *well* in *he treats us well* seems to be an argument, although it’s a manner expression. Yet other criteria are based on argument indexation and flagging: an argument with oblique case marking does not leave agreement markers on the main verb, while an argument with structural case (nominative, accusative) does if the language has agreement marking, and the latter is recognized as a core argument. Following this standard, many so-called oblique arguments, like *this* in *I think [of this]*, would be classified as peripheral, although they are clearly licensed by the lexical entry of the verb. These criteria correlate with each other but in a non-deterministic way.

Because Mandarin has no verb agreement, only the first two criteria can be used, and the problems listed above all occur. The status of comitative 跟你 in (5) is not so clear, for instance: it is fairly low in Fig. 1.1, but it is not obligatory. We also note that reordering of peripheral arguments is possible, but mixing them with TAME markers sounds problematic to say the least (8). Note that fronting of the comitative to a higher position is possible (§ 1.4).

- (8) a. 我明天可能能在我的办公室跟你讨论一下 (=5)
 b. 我明天可能能跟你在我的办公室讨论一下
 c. ??我明天可能跟你能在我的办公室讨论一下

⁴Note that English adverbs like *allegedly* follow the same generalization: we have e.g. *he allegedly possibly did this* but never *he possibly allegedly did this*. See Cinque (1999).

1.2.4 The structure of the core verb phrase

In the surface form, the core VP contains the core arguments and the **verbal complex**. The verbal complex involves aspectual marking (§ 1.2.3.1), verb derivation and verbal complements (see below); lexical aspect is not grammatically marked but is important.

- (9) 我 [做完 了]_{verbal complex} 作业
1 do finish ASP homework
'I have finished the homework.'

1.2.4.1 The verbal complex

The verbal complex contains three

Mandarin has lots of clausal complements that are not prototypically arguments, known as 补语 in Chinese linguistic community.⁵ This is a rather heterogeneous category, its boundary (expectedly) being somewhat unclear; it includes verbal complements or in other words complex predicates, complement clauses, and oblique arguments.

1.2.4.2 Verb-final constructions

Mandarin has two types of constructions in which the main verb appears at the final: the DISPOSAL construction, and the so-called PASSIVE construction, also known as the *ba*-construction and the *bei*-construction.

1.2.4.3 Transitivity

Verb frames in Mandarin can be divided into the DO class (about actions), the BECOME class (about changes of states), the CAUSE-BECOME class (about something causing a state to change), and the purely stative BE class.

Just like the case in English, DO verb frames often allow S/A-ambivalence, where the P argument of a transitive verb frame (i.e. the more internal, patient-like argument) can be removed, leaving only the subject (10).

- (10) a. 他喜欢玩
b. 他喜欢玩玩具

On the other hand, we observe regular alternations between BECOME and CAUSE-BECOME verbs, and hence S/P-ambivalence (11). It should be noted that not all BECOME verbs can receive a CAUSE-BECOME verb frame (12).

- (11) a. 茶泡好了
b. 我泡好茶了

- (12) a. 这只猫死了
b. *坏人死了这只猫

⁵The term 补语 literally means 'complementation speech', and is therefore often translated as *complement*. In this note I use the term *complement* to refer to grammatical constituents that are somehow more closely related to the lexical head, and I choose the (somehow tedious but explicit) term *non-argument complement*.

The verb frames naturally have correlations with the lexical aspect of the clause: change-of-state clauses are naturally telic and often cannot be in the progressive aspect (13). But counterexamples exist (14).

(13) *这只猫正在死

(14) 我正在泡茶

1.2.4.4 Verb-particle constructions

Just like English, Mandarin has directional and resultative particles in the argument structure.

1.3 Parts of speech

After a survey of the grammatical system of Mandarin Chinese, we examine what the lexicon has to feed into the grammar.

1.4 Subject being topicalized

(15) 我明天跟你可能能在我的办公室讨论一下

1.5 (Absence of) dangling topics

Some people, like Zhu (2009, § 7.1), equate *subject* with *topic* in Mandarin grammar. Some (especially those from the functional-typological tradition) go further and assert that “the notion of the subject (as the position of the most agentive argument) isn’t grammaticalized in Mandarin Chinese”, and therefore the topic-comment construction is construed as simply the syntactic coding of aboutness, and this base-generated and syntactically unconstrained topic is called a “dangling topic”. This view is rejected in this note, because such accounts usually end up in severe overgeneration. Here I briefly summarize Shi (2000)’s argumentation.

1.5.1 Type 1: Idiomatic phrasal predicate looking like a comment

In the first type of “dangling topic”, it’s impossible for any NP in the comment to be syntactically related to the topic (16, 17). Such cases however should be analyzed as instances of the subject-predicate construction, where the predicate is a dephrasalized clause.

We notice that in such examples, the “comment” often has already undergone various degrees of fossilization. Changing the comment usually makes the sentences much less felicitous (18), at best highly marked. This is strange if the attested examples are topic-comment constructions, but makes sense if dephrasalization is needed to put the clause 大鱼吃小鱼 etc. to the “comment” position.

Thus, in (16) and (17), the so-called topic is an ordinary subject, and the so-called comment is a predicate. The meaning of the result of dephrasalization may be compared with the English colloquial *I was like, ...* construction.

- (16) 他们[大鱼吃小鱼] (， 厮杀成一片)
 (17) 他们[你看看我我看看你]
 (18) a. *他们小鱼咬大鱼
 b. *他们虾米啃泥底

1.5.2 Type 2: Quantificational adverbial looking like the inner subject

The second type of “dangling topic” is like (19). A topic-comment analysis of (19)

- (19) 他们 谁 都 不 怕
 3pl who even NEG fear
 ‘They don’t fear anyone.’

1.5.3 Type 3: Ellipsis leaving a subject and one predicate

Some people accept (20). Here the NP 那所房子 definitely doesn’t come from the words following it, and is therefore recognized as a topic by some (TODO: ref). Note, however, that 幸亏 serves as a clause linker outside (20): (21) is a demonstration of the 幸亏……不然…… linking construction, and we also have its topicalized version (22). (TODO: whether this is parenthesis) We also know in a clause linking construction, often one clause can be omitted in the utterance because it’s content can be easily inferred (TODO: ref). So now the origin of (20) is clear: We can get it by omitting the second clause in the comment part of (22). Indeed, if we replace 幸亏 by anything that is adverbial but not a clause linker, the resulting sentence – which now contains a real dangling topic – is not grammatical.

- (20) % 那 所 房 子 幸 亏 没 下 雪
 that CL house fortunate NEG snow
 ‘For that house, fortunately it didn’t snow (or otherwise something bad would happen).’
 (21) [幸亏] 去 年 没 下 雪 ， [不然] 那 所 房 子 早 就 塌 了
 fortunate last.year NEG snow otherwise that CL house already collapse SFP
 ‘Fortunately it didn’t snow last year, or otherwise that house has already collapsed.’
 (22) [那 所 房 子]_{topic} [幸 亏 去 年 没 下 雪 ， 不 然 早 就
 that CL house fortunate last.year NEG snow otherwise already
 塌 了]_{comment}
 collapse SFP

1.5.4 Type 4: Extraction from prepositional adverbials

(??) in § ?? is sometimes regarded as an instance of the dangling topic construction. However, as is shown in § ??, it may just be from topicalization of an NP in an adverbial, with the preposition (and/or the locative particle) removed.

1.5.5 Type 5: Nominal predicate

(23) 这种青菜一斤三十块钱

1.5.6 Type 6: Locational adverbial mistaken for the subject

(24) % 物价 纽约 最 贵
price New.York most expensive
'The price in New York is the most expensive.'

1.5.7 Tentative conclusion

The conclusion is all topics in Chinese are closely linked to a position in the comment, be it a core argument position or a peripheral one. So the notion of dangling topics is to be rejected in Mandarin grammar, and we can always recover the “canonical” i.e. non-topic-comment clause from a topic-comment construction. After this, if the canonical clause can be divided into an NP or a complement clause and a verbal constituent following it, we can uncontroversially say the first is the subject while the second is the predicate. (TODO: predicate def) So equating the subject with the topic is also wrong.

It's possible to find the semantic role of the subject isn't agentive; in this case I assert there is a valency changing mechanism here.

Box 1.2: What to expect when people talk about the subject or the topic

Unfortunately, despite the syntactic tests presented above, there are still many people – even many native speakers – promoting the idea that the Mandarin topic has nothing different with the subject. Here is a list of TODO: ref