

Mandarin Chinese notes

Jinyuan Wu

February 23, 2023

Chapter 1

Introduction

1.1 The language and the speakers

Mandarin Chinese is a predominant language in the world, belonging to the Sinitic family. Indeed,

1.2 Previous studies and theoretical orientation of this work

1.2.1 Structuralist generative grammars

There are already sufficient works concerning the grammar of Mandarin Chinese. The tradition used in colleges (and occasionally high schools) is largely structuralist à la Bloomfield: A clause is divided into a topic and a comment, and the comment is divided into a subject and a predicate, and the predicate is divided into a predicator and an object, etc. Examples of works in this tradition usually have names like 现代汉语 ‘Modern Chinese’ (王理嘉 et al. 2004). This tradition is still seen in many contemporary grammars, like Huddleston and Pullum (2002). This tradition is largely coherent with the generative tradition, and indeed there are books which are generative in essence but organized in the traditional and structuralist framework (邓思颖 2010).

1.2.2 Teaching materials

Teaching materials of Mandarin Chinese are also largely influenced by the structuralist tradition. TODO

1.2.3 Mandarin in the functional-typological tradition

Mandarin also gains much attention in the functional-typological tradition. Li and Thompson (1989) is a “functional” reference grammar of Mandarin, TODO The reason for Mandarin’s popularity seems to be the fact that it breaks many previous typological generalizations about isolating languages and constituent orders (Paul 2014, chap. 8).

1.2.4 Theoretical commitment of this work

This notes attempts to reconcile the several approaches in previous researches. It’s my belief that the differences between the generative tradition, the traditional structuralist school and a large part of the so-called functionalist (or “Basic Linguistic Theory” (Dixon 2009)) are mainly notational from a practical perspective. Below, by the capitalized Generativism, I mean a mixture of Minimalism, Distributed Morphology and Cartography (but with less flavor of Antisymmetry), which I mean are instructive on grammar description. This leaves out Lexicalist traditions, due to the limit of space. Indeed, what I want to do here is to incorporate the new perspectives in generative syntax, such as Paul (2014) and Paul (2008), into the structuralist tradition in an accessible and typology-informed way.

Since modern generative syntax contains lots of hidden functional heads, the notion of, say, a DP which is the specifier of T, is to be replaced by an NP filling a subject position in a surface-oriented structuralist constituency analysis. Both “NP” and “subject” should be labeled on a sub-syntactic tree in the structuralist tradition, while in generative syntax, the label “subject” is a secondary concept: it’s an abbreviation of SpecTP (or in an even more fine-grained way, SpecSubjP or SpecNomP). So functional heads can be replaced by syntactic function labels. On the other hand, in Distributed Morphology we

have roots, which reside at the center of an extended verbal or nominal projection (NP-NumP-DP, or vP-TP-CP), and they are recognized as *heads* (in this note referred to as **lexical heads**) in traditional structuralism. This unifies the notation of Huddleston and Pullum (2002), Chao (1965), 朱德熙 (2009) in the traditional structuralist perspective and Generativism.

The Basic Linguistic Theory (BLT) approach (Dixon 2009), or the grammar writing approach accepted in modern descriptive linguistics and typology, is based on *dependency relations* on the other hand. It's also possible to formulate generative grammar in terms of dependency relations, so it can be expected that the BLT approach is still largely equivalent with the grammatical complexity of generative grammar. BLT only recognizes two major types of constituents: NPs and clauses, which are essentially *domains* or *fields* in generative syntax: The former is the DP domain while the latter is the vP-TP-CP domain. The complicated binary constituency tree is replaced by a flat tree with lots of dependency relations labeled inside, containing the same amount of information. (We may also say the Basic Linguistic Theory's standard of constituency is "being a relatively independent construction". Then the equivalence between this version of mild constructivism and minimalism is clear (Trotzke 2020).) Note that the Basic Linguistic Theory approach still (although quite implicitly) admits that there is a rank of "closeness" or "height" among dependency relations: the A argument seems to be somehow higher than the O argument, etc., and this information is conveyed by the fine-grained constituency relations in Generativism.

The focus on dependency relation in BLT also leads to a notational difference on what is a constituent. A sequence like *has been exploring* is not a constituent in the sense of the structuralist constituency test in grammars like Huddleston and Pullum (2002) and generativism, but are still recognized as a unit in BLT because its parts always appear together in the surface-oriented analysis. This what is Dixon (2009, p. 109) defines as a *verb phrase*, which excludes the object. From the constituency-based point, such a "phrase" is usually a larger functional domain minus a smaller functional domain; here it's the functional projection below the subject – the verb phrase in Huddleston and Pullum (2002) – minus the object. This definition is meaningful because syntax is cyclic and the object is a DP and therefore is a phase, and above it is another phase, and although the introduction of the subject (by, say, a Nom head which may contain the nominative case) doesn't seal a phase, it still changes the properties of the syntactic tree, so what happen above the object and below the subject – the TAME functional heads, the verb root, etc. – are tightly linked to each other (note that here we are just translating between the constituency-based analysis and the dependency-based analysis), so from a dependency relation-oriented analysis, the sequence *has been exploring* does qualify as a phrase; in a constituency grammar we recognize this sequence has a status but we don't call it a phrase or a constituent. This difference is notational but may be confusing.

A further difference between the Basic Linguistic Theory and Generativism (as well as the structuralist tradition) is the former is claimed to be semantic-based. But first, of course it's possible to have a meaning-first version of generative syntax, and second, there seems to be a "gluing" layer between pure semantics and the phonetic realization and can't be equated with either of the two: the semantics of complement-taking verbs may be coded as a complement clause construction, a relative clause construction (compare *I see a man running* and *I see a running man*), and superficially similar utterances may have different "structures". This is recognized by Dixon, who distinguishes the "prototypical" coding strategy of a semantic concept and other strategies. So there is also no substantial disagreement here.

The Generativism I take here doesn't emphasize on wordhood as a universal concept, and indeed this is what I want here: A word is simply a mini phrase (in the BLT sense), either the realization of a mini constituency tree, or the realization of a span of functional heads and possibly the lexical head or in other words the root. The controversy about what is a word in Mandarin has been around for decades, which I believe is due to the desire to find *the* word as a universal unit, without thinking of the tenet of Generativism that phonetic realization doesn't always transparently reflect in the syntax proper, and that what is universal is likely to be prototypes of functional heads and how they are arranged together; this means we may have categorizer phrases like nP or vP, which can be recognized as minimal words, but whether things like compound words are recognized as words or phrases depends more on grammatical traditions and external factors like prosody.

1.3 Plan of the book

1.3.1 Origin of data and how to represent them

I will use less structuralist constituency trees compared with structuralist grammars. That's to say, for example, the "serial verb construction" is not divided into two verb phrases following rigidly the structuralist tenets, but is analyzed with a flat-tree structure instead, following BLT. But I will also do lots of in-depth morphosyntactic tests, instead of just staring at the surface realization, as many typologists may do.

1.4 Remarkable features of Mandarin

1.4.1 Lack of morphology?

Despite of lack of inflection and lack of contextual alternation of morphemes, Chinese does have some local and syntactically unmotivated operations which are just like morphophonological rules, although they don't necessarily operate on phrases.

An example of this is the verb copying phenomenon, as in 看了一会书 (compare 看书了一会); 看书 'to read' (intransitive; lit. 'to read books') is a fossilized verb-object structure and this verb-object structure may still have synchronic effects. More radical examples however also exist, like %体了一堂操, which is likely to be linked to [[体操]_{noun-as-verb}了[一堂]_{time object}]_{VP}. In casual speech, verbs borrowed from other languages may also be split and the two fragments of the verb then surround the semi-object (1).

(1) 我 [debug]_{topic:VP} [de不出来]_{predicate:VP} [啊]_{SFP}

This phenomenon – the verb being split and the time semi-object getting embedded into the verb – looks just like infixing, although here this infixing operation targets a VP instead of a smaller unit. This justifies the assumption taken at the end of § 1.2.4 that there is no clear boundary between words and phrases and therefore syntax and morphology: It's possible for a phrase to undergo rearrangement without clear syntax motivation that usually happens within a word.

1.4.2 The overwhelming influence of prosody

One distinct feature of Mandarin is its morphosyntax relies strongly on *prosody* (冯胜利 2000). Other components in phonology, strikingly, doesn't have much influence on Chinese morphosyntax, and it will be largely skipped in this note.

1.4.3 The so-called serial verb constructions

It's often said Mandarin is a serializing (i.e. with serial verb constructions) language. A closer look, however, reveals this is not the case: These constructions are either adverbial clause constructions or complement clause constructions, or maybe certain kind of light verb constructions (§ 8.3.1). The internal heterogeneity renders the term *serial verb construction* useless.

1.4.4 Ionized verbs

Some verbal units – commonly recognized as verbs, i.e. grammatical words – in Mandarin may be ionized into two parts, with a constituent residing between the two. This involves two mechanisms: the first is TODO: 念佛-like, in which a verb phrase is fossilized, similar to English *in case* or *by definition*, but its object can be extended into a complex NP, and the second seems to be a morphophonological device, by which even a verb without synchronically analyzable inner structure is teared into two parts and a clausal dependent is inserted between the two (TODO: ref). Despite its striking properties for English speakers, ingredients of this phenomenon are all well attested cross-linguistically.

Chapter 2

Phonology and the writing system

2.1 Prosody

By paying attention to stops in Chinese utterances, it can be found that phonological words exist and they are mostly defined by the prosody structure. In the rest of this note, the term *prosodic word* and *phonological word* will be used interchangeably. The prosody structure is about how stress is assigned to phonological constituents. Assigning a prosodic structure is like condensation and clustering: something is merged with something adjacent, and the result is merged with something adjacent else. When two phonological constituents are merged together, one of them is considered heavier than the other. If heaviness is to have a simple relation with the length of a phonological constituent, then usually the more a phonological constituent is, the heavier it is. This is consistent with the condensation picture of prosodic segmentation. Suppose a prosodic constituent attracts a syllable and merges with it. The latter is not an independent phonological constituent and cannot be heavy, so the former is the heavier one and the latter is the lighter one in the larger prosodic constituent.

The smallest unit of prosody structure is a prosodic word. The simplest prosodic word is the disyllabic foot, which contains two adjacent syllables in the case of Chinese. (It can be made by two moras in other languages.) One is assigned stress and is therefore heavier than the other. Trisyllabic prosodic words also exist in Chinese, though they are highly limited. Most of which are borrowed words (e.g. 加拿大 ‘Canada’) or words formed by coordinating three morphemes (e.g. 数理化 ‘math, physics, and chemistry’). They can also be regarded as feet (冯胜利 2000, § 2.2).

Longer morphosyntactic units are inevitably broke into smaller disyllabic or trisyllabic prosodic words in their prosodic structures, often regardless of their morphosyntactic structure: 加利福尼亚 may be segmented into 加利|福尼亚, although the word contains only one morpheme. In 副总经理, we have two prosodic words, 副总 and 经理, while the morphosyntactic structure of the word is [[副][总[经理]]] (TODO: ref to what the structures are). This is similar to the case in English and Latin poems, where the prosody arrangement of sentences does not have to respect word boundaries: *arma vi|rumque ca|no*. It’s however also possible that a prosodic word has morphosyntactic significance (§ 3.1.2.2).

Prosody is able to see the constituency structure and prosodic constraints are important in Mandarin grammar. Some prosodic rules pertaining to the constituency tree guide and limit the assignment of relative heaviness and lightness. In Chinese, prosodic segmentation is done strictly left-to-right in each NP, and then the NPs together with verbal constituents are used as the input of prosodic segmentation of clauses. Certain forms are therefore ruled out (§ 8.2), not by morphosyntactic reasons but for prosodic reasons.

2.2 Chinese characters

Chapter 3

Parts of speech

3.1 From morphemes to clauses: levels of units in Mandarin morphosyntax

3.1.1 Morphemes

A morpheme is a minimal unit in grammatical analysis. The meaning of the term has some ambiguity: sometimes, a grammatical item has a clearly analyzable inner structure but the structure no longer has any morphosyntactic significance (§ 3.1.1.2, § 3.1.1.3). In this case, we say the item is *synchronically* a morpheme.

I start the introduction of Mandarin morphosyntax with morphemes for good reasons. Although in most language documentation projects, words – whatever this term mean (§ 3.1.2) – are the smallest unit appearing in the dictionary, this is not the case with Mandarin: the standard practice of lexicography is to record Chinese characters and their “meanings”; in linguistic terms, an entry of a Chinese character includes the follows: its pronunciation(s); historical or synchronic monosyllabic morphemes that may be represented by that character and polysyllabic morphemes containing that character, which may have different pronunciations; (§ 3.1.1.1, § 3.1.1.3, § 3.1.1.2); grammatical words that are made up regularly using one of the morphemes listed above and have already gained a stable meaning. Thus, for someone who wants to do a thorough grammatical analysis of an utterance, finding the morphemes – instead of grammatical words recognized in § 3.1.2 – is the first step.

3.1.1.1 Primitive morphemes

Most native primitive morphemes are monosyllabic, examples of which include 红, 大, 你, etc. (§ 3.1.2.1). Polysyllabic primitive morphemes, like 葡萄, 巧克力, 哥斯达黎加 etc., are mostly borrowed words in different historical stages. There do exist seemingly native polysyllabic primitive morphemes, like TODO: historical analysis of 轱辘, etc.

Only a subset of the monosyllabic morphemes are able to serve as grammatical words. Some monosyllabic morphemes only appear in in-word slots (like the modifier slot in (3b)), and they are unable to serve as words. Many of them are historically free morphemes but have become obsolete in contemporary usage. The morpheme 观 ‘observe’, for example, still exists in 观鸟 ‘bird watching’ but never appears as a single verb, nor does it undergo delimitative reduplication of verbs (1).

- (1) a. *我要去观观那些鸟
b. 我要去看看那些鸟

An overwhelmingly portion of polysyllabic morphemes are able to constitute one-morpheme word and also one-word phrase. There exist however a small number of polysyllabic morphemes that seem to be unable to be words themselves. The root 日耳曼 (from *German*) appears in compound nouns like 日耳曼人 or 日耳曼血统, but almost never appears as a noun itself; similarly, 达达 in 达达主义 also never appears as a word itself.

3.1.1.2 Fossilization

A lot of words have internal structures parallel to those observed in syntax (朱德熙 2009, § 2.6), but the structures have already completely fossilized, so they may be synchronically regarded as containing

only one morpheme. A completely fossilized structure may be historically created with an obsolete syntactic device, or be not in the expected part of speech inferred from its etymological (i.e. it appears in syntactic environments that are not expected for its inner structure).

The word 例如 ‘an example is like ...’ is a connective adverb, but it has a subject-predicate with a gap.

3.1.1.3 Abbreviation

There is a strong tendency to make the abbreviation a prosodic word. The abbreviation of a binary-branching structure usually consists of the first syllables of its two immediate constituents, regardless of the inner structures of the two immediate constituents. Thus [副 [总经理]] ‘vice general manager’ is usually abbreviated as 副总, and 总工程师 is usually abbreviated as 总工. It’s possible that an abbreviation replaces the original word completely. 空调 is historically the abbreviation of 空气调节器, the word-by-word translation of *air conditioner*, but the latter is no longer in active use.

Trisyllabic prosodic words do exist in Mandarin, and trisyllabic abbreviations also exist. The majority of them are three-morpheme coordination structures like 数理化, which is the abbreviation of 数学物理化学 ‘math, physics, and chemistry’.

Box 3.1: No generative rule in the lexicon

Those insisting on a universal word-phrase distinction may say “fossilized structures are assembled in the lexicon before syntax”. The position of this note, however, is if something is assembled synchronically, then it has to have something to do with syntax: syntax is the only productive engine. If a morphological device is completely invisible to the rest of the grammar, it is likely to have lost productivity and becomes historical; its products are therefore synchronically morphemes, instead of words with inner structures.

3.1.1.4 Function items and semi-function items

Another type of morphemes is the type of function items,¹ like function words, inflection suffixes (which are lacking in Mandarin). Distinguishing between words and phrases means to classify morphosyntactic units according to their possible internal grammatical relations and categories. Thus, we distinguish between NPs and compound nouns, for they possess different internal structures (§ 3.3.4). A function item, on the other hand, is a *label* of a grammatical category and has no inner structure. Thus, defining whether something is a function word or a function morpheme is purely based on *external* factors: if something appears in a morphosyntactic unit commonly referred to as a phrase, than it’s a function word, and if it appears in a morphosyntactic unit commonly referred to as a grammatical word, than it’s a bound function morpheme. But there seems to be no particularly strong correlation between whether a function item is grammatically a function word and whether it is phonologically a word: a suffix in a long grammatical word can be a phonological word (TODO: Basque example?), while a function item, like the Latin *-que*, that is an immediate constituent of a phrase may be glued to another phonological word. In general, the meaning of *function word* is not well-defined, and an accurate description of a function item inevitably involves both of its phonological status and its grammatical status. That’s why I refrain from discussing wordhood of function items in the following sections.

Monosyllabic location words, for example, appears in NPs and therefore should be analyzed as words; they however never constituent one-word NPs themselves (§ 3.5). Whether they are words therefore becomes a mystery, and this mystery is merely a wrongly asked question.

Certain borrowed affixes may be unable to serve as words in certain periods. As times goes by, however, they gradually become free morphemes. If clause linking markers like 之所以 and 是因为 are recognized as single morpheme words, then they may be included into the non-prosodic simple word blob and however be unable to serve as phrases. These markers, however, never appear in other places, and their exact status is of no descriptive and comparative interest.

¹Here the term *grammatical word* means a word defined in the grammar, not phonology; a *function item* means an item that is not lexical and therefore is a part of the grammar. The adjective *grammatical* means being related to grammar and it has two explanations: being defined in grammar and being contained by the grammar; In this note, I use the term *grammatical* to refer to the first meaning, and *function* to refer to the second meaning.

Box 3.2: On some confusing notions of morpheme classification

Many publications on Mandarin based on structuralism à la Bloomfield have some outdated, confusing, and not really necessary notions on classifying morphemes. 朱德熙 (2009, p. 16) claims that free morphemes don't have fixed positions, while bound morphemes sometimes have fixed positions. Cross-linguistically, this is simply wrong: in SOV languages like Japanese, the position of the verb is predominantly after all clausal dependents and before inflectional endings, but we would all agree that verb roots are prototypical free morphemes. On the other hand, bound morphemes like derivational suffixes do have fixed positions, but that simply comes from the fact that they have fixed positions in the syntactic structure – and indeed so-called free morphemes can only occupy a limited number of positions in the syntactic structure (lexical head position, compounding attributive position, etc.) and this may lead to a constituent order effect, as is shown in the above example of Japanese.

3.1.2 The existence of words as mini-constituents

A question causing endless controversy and confusion is “what is a word”. BLT spends a whole chapter (chap. 10) on this topic. It is often said that Chinese is “character-based” or to be precise, “monosyllabic morpheme-based”, with no level of grammatical words. This claim is factually flawed, since in Chinese, there *are* distinction between productive morphemes and words. What should be noted are the split between phonological words and grammatical words and the subtleties concerning word-phrase distinction. These are introduced in the following sections.

3.1.2.1 Monosyllabic words

Some words TODO: 蛇、雨

3.1.2.2 Linking prosodic words to morphosyntactic words

We expect a grammatical word to be a mini-constituent, which is therefore easily subject to conventionalization, and has limited interaction with the syntactic environment (although definitely not absolutely no interaction – even in English, we have things like *pre- and post-processing*, in which a phrasal structure – coordination – interacts with the inner structures of two bare nouns, and therefore a clear boundary between word or phrase or between morphology and syntax is in principle impossible (§ 3.1.3)).

A phonological word (§ 2.1) may be a single-morpheme with a definite part of speech tag (§ 3.1.1), like 幽默 and relatively rare trisyllabic cases like 加拿大, or a mini-constituent (§ 3.1.2.3), like 白菜 or 种树 (TODO: ref), or a sequence that according to the constituency-based analysis is not a grammatical unit but does package highly related items (§ 3.1.2.4), like 交给 (TODO: ref: verbal complex). But of course it's possible that a phonological word has no morphosyntactic significance at all.

This then raises the question whether all prosodic words that are also morphosyntactic units are small enough, or some of them are actually phrases. This question can never be answered satisfactorily (§ 3.1.3). The observed fact that the morphosyntactic interpretations of phonological words are all small units and either have conventionalized meanings or are unable to be extended infinitely urges me to use the term *grammatical words* for them, although the definition of the latter is admittedly subjective, although the hierarchical structure of the grammar reflected by the definition(s) is not. Indeed, sometimes it's the phonological wordhood that settles down whether a sequence is recognized as a word in everyday use or not (§ 3.1.2.4).

There are also non-prosodic words in Mandarin, including polysyllabic words like 哥斯达黎加 (§ 3.1.1.1) and complex non-prosodic words like 总工程师 (§ 3.1.2.5). Thus I call prosodic words with morphosyntactic significance **morphosyntactic prosodic words**.

3.1.2.3 Disyllabic mini-constituent words

There are disyllabic units in Chinese that have conventionalized meanings and its inner structure is invisible to any other morphosyntactic rules (§ 3.3.2). Examples of them include 念佛, 军训, 体操 etc.

Many of these disyllabic word are made up by two morphemes with a synchronically available device. Thus, the two morphemes may be analyzed as both grammatical words and morphemes. Then

it's possible that the morphosyntactic prosodic word can be extended into a phrase by inserting more phrasal dependents or by adjoining modifiers to one of the two morphemes. This is usually not the case for nouns, but is indeed true for verbs (§ 3.4.1), raising questions concerning their wordhood. But note that similar processes are possible even for disyllabic verbs without analyzable inner structures, which are surely grammatical words (§ 8.4).

3.1.2.4 Prosodic words with questionable constituenthood

The standard for a constituent varies, and so does the definition of a mini-constituent. The English form *explore-ed* still needs to take an object, and some syntactic and semantic analysis actually implies that the tense/aspect categories are somehow higher than the object. Thus it can be recognized as a unit – a constituent – because its inner components have strong interlink in dependency relations (§ 1.2.4). In principle, we can also analyze it as *[-ed [explore object]_{argument structure}]_{TAME marking}*; in practice we refrain from doing so (except in fully theoretical works like Distributed Morphology), because *explored* is a *phonological* word (§ 3.1.2.2), and this urges us to recognize it also as a morphosyntactic unit (essentially, by redefining what is a unit). The longer sequence *have explored* has a parallel structure, but is not a phonological unit, and therefore some grammars say it's a verb phrase, while others say it's not a phrase. (TODO: ref) Both analyses, of course, are well-justified – which one is to be used depends on the focus of the grammar, and by alternating the terminology the two analyses are equivalent (§ 1.2.4). The definition of the term *word* therefore in principle carries no assertion regarding the nature of the language in question; we do need to choose the definition carefully enough, though, to avoid confusing people.

In Mandarin, although prototypical inflection is absent, in verbal complement constructions and the aspect system, we indeed can see something similar to *explored* mentioned above, the inner parts of which have strong dependency but do not form a constituent in the most strict sense. What is the status of 爬上 in 他笨手笨脚地爬上信号塔? A word (created by a productive verb compounding rule), a phrase (a verb-complement structure), or just a word sequence without structural significance? Here I follow the opinion in (冯胜利 2000, p. 86) and (Tham 2015) and call it a word, because a sequence like 爬上 is extended in a highly limited way (the only possibilities being 爬得上 and 爬不上), while phrases, in principle, can be extended infinitely, and it is also a prosodic word. This goes against the analysis in (朱德熙 2009, § 1.2.7).

3.1.2.5 Non-prosodic complex words

Certain grammatical relations seem to be not a part of NPs and clauses, again highlighting the necessity to introduce a smaller level of constituency (§ 3.3.4, TODO: verb), commonly known as grammatical words. The structures listed above in principle can be extended without an upper bound, and therefore they are not monosyllabic and are not prosodic words.

Just like the case of morphosyntactic prosodic words, these **non-prosodic complex grammatical words** may be created by synchronic morphosyntactic rules, or they may be fossilized or have no internal structure.

Compared to prosodic words, non-prosodic complex words are less “active” in syntax: splitting them is possible in certain cases but is much less frequent. This may be a result of pragmatics: complex words are created to cover a meaning that needs some explanation, and once a complex word is well-accepted, its form and meaning soon gets fixed (because people will not burden themselves), and fossilization occurs rapidly. The term 美利坚合众国 has an analyzable internal structure, but it has already gained a fixed meaning and its parts are never taken out, despite both 美利坚 and 合众国 can serve as grammatical words.

3.1.3 The blurred boundary between words and phrases

3.1.3.1 Fossilized phrasal structures

One peculiar feature of Chinese is grammatical relations in morphology often have syntactic counterparts, with the same constituent order (§ 3.4.1). Some still have synchronically analyzable inner structures, as in 念佛, while others have already lose inner structures (as in [关心]她, where 关系 has collapsed into a single morpheme synchronically).

3.1.3.2 Splitting of word

One thing that makes the concept of word somehow puzzling is even words *without* synchronically morphosyntactic internal structures can sometimes be split and extended with phrasal dependents injected, though not everyone will accept such usages. The only attested case is in the clause structure, where the main verb may be split or ionized into two parts (§ 8.4), even when the verb does *not* have a inner verb-object structure (§ 3.4.1).

3.1.4 Phrases

3.1.4.1 Noun phrase

Nouns are able to be lexical heads of NPs, including non-prosodic simple words like 哥斯达黎加, as well as non-prosodic complex words. One-word phrases are always possible, and there is almost no attested counterexample.

3.1.4.2 The verb phrase

Transitive verbs can regularly fill argument slots and thus are able to be used as one-word phrases, though they themselves are not sufficient to build one-word predicate VPs. TODO: really??? Certain grammatical words, like some verbal complement structures recognized as grammatical words (§ ??), are not phrases on their own.

- (2)
- a. [看书]_{subject:verb} 是一件有趣的事情
 - b. *[走进]_{subject:verb} 意味着您已经同意了我们的服务条款
 - c. [走进这个建筑]_{subject:VP} 意味着您已经同意了我们的服务条款

3.1.5 Types of clauses

3.1.6 The structure of Mandarin grammar

After establishing the status of *words* in Mandarin grammar, we can now discuss the overall architecture of Mandarin grammar in a more disciplined way. Fig. 3.1 summarizes the organization of Chinese lexicon as well as how larger units are built from lexical items. Overlapping of blobs means “having the same form”. Thus, the blob representing monosyllabic words is completely in the blob of monosyllabic morphemes. The same is for the relation between non-prosodic simple words (which are neither monosyllabic nor disyllabic) and polysyllabic morphemes. Red arrows mean synchronic morphosyntactic devices, while orange arrows means historical evolution, like grammaticalization and/or fossilization,.

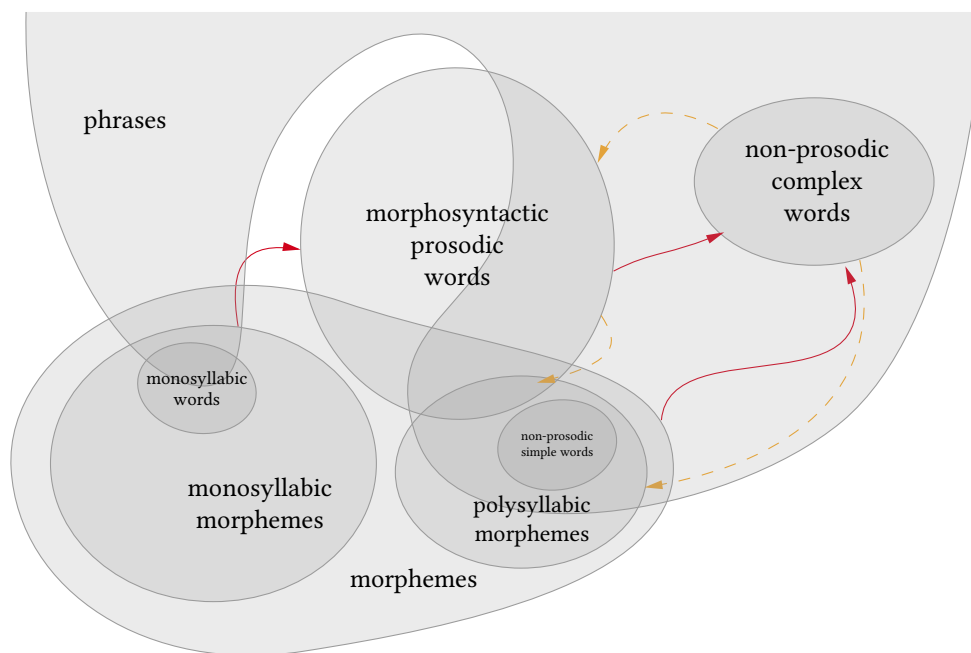


Figure 3.1: From morphemes to phrases.

The sub-phrasal units in Fig. 3.1 are represented in Table 3.1. The +? symbols in the “morpheme” column mean possible fossilization, corresponding to the lower two orange arrows in Fig. 3.1. The upper orange arrow in Fig. 3.1 means word creation by abbreviation. The +? symbol appearing in the “phrase” column means some authors may agree that 走进 is a constituent and therefore they may say it’s a phrase (§ 3.1.2.4).

3.1.7 Perception of basic units in Mandarin

In English there is a group of clearly defined and largely homogenous morphosyntactic units lying between morphemes and phrases (which means they can be neither morphemes nor phrases, and can be constructed by the former and can be used to build the latter), which is just the grammatical word. They are recognized as basic units by society (i.e. outside of the linguistic community): the length of an article is measured by counting words, not letters, for example.

In Chinese, however, despite the fact that we can still identify grammatical words, that grammatical words have much larger overlapping with morphemes and phrases means Chinese characters – representation of syllables in polysyllabic morphemes and rough representation of monosyllabic morphemes – are recognized as basic units of the language. This may be the deriving force for some linguists (who are too eager to “find diversity”) to reject the existence of grammatical words in a hurry, although I’ve already shown this is not the case.

Still, native speakers generally confirm that there is something between the morpheme and the phrase, which people call 词. In everyday speech, non-prosodic complex words have limited use, and monosyllabic words are rarely used because their formality (§ 3.1.1.1). Thus, from Fig. 3.1, we find it’s the class of prosodic morphosyntactic words that plays the role of “words” as a level between morphemes and phrases, as well as intermediate units between morphemes and complex words. Prosodic morphosyntactic words are also usually the final destine of abbreviation (§ 3.1.1.3). These morphosyntactic prosodic words are therefore the nature response when native speakers without much linguistic training come up with intuitively when talking about “words” or 词, i.e. intermediate building blocks.

The fuss around wordhood in Chinese arises from the split of several definitions of the term *word*: wordhood defined by morphosyntactic test include is wider than the coverage of morphosyntactic prosodic words, and the range of morphosyntactic words indeed has too much overlapping with morphemes and phrases.

Table 3.1: Sub-phrasal units

index	length	morpheme	word	phrase	example	comment
1	monosyllable	+	-	-	竟, 语	usually from Classical Chinese
2		+	-	-	性 in 等价性, 前 in 窗前	suffixes, location words, and other semi-grammatical items
3		+	+	+	蛇, 猫	usually from Classical Chinese
4	disyllable	+	-	-	达 in 达达主义	usually borrowed terms, rare
5		+	+	-	觊觎, 葡萄, 空调	from Classical Chinese, borrowed items, abbreviations, and deep fossilization of type 6 and 7
6		-(/+(?))	+	-(/+(?))	走进, 来到	constituents in BLT only; made of monosyllables
7	trisyllable or larger	-(/+(?))	+	+	念佛, 看书, 吃饭	mini-constituents; made of monosyllables
8		+	-	-	日耳曼	usually borrowed terms, rare
9		+	+	+	哥斯达黎加, 滑铁卢, 数理化	usually borrowed terms or abbreviations
10		-(/+(?))	+	+	副总经理, 总干事	complex word, made of monosyllables and disyllables

3.2 Overview of classification criteria

3.2.1 Word class labels: noun, verb, adjectives, etc.

3.2.1.1 The nominal-verbal division

Lexical words in Chinese can be roughly divided into nominal ones and verbal ones, or in the Chinese terms, 体词 and 谓词. The prototypical role of nominal words is to fill predicate slots (or to be more precise, to head a phrase that fills an argument slot). Nominal words rarely appear in the verbal complex, though for stylistic purposes, they sometimes do. Verbal words prototypically appear in the verbal complex (chap. 5), but many of them – and clauses without any morphological marking – can regularly appear in argument slots (朱德熙 2009, § 3.5).

The fact that verbal categories can fill argument slots or in colloquial words “be used as nouns” urges some to put the verbal categories under the nominal categories, so thus there is only one mega lexical category in Chinese: the nominal category or the Noun. The analysis adopted here does not aim to organize lexical categories in a binary branching classification tree, so the ordinary nominal-verbal distinction is maintained: verbs being able to fill argument slots is not typologically rare, actually, and this shared feature itself does not bring nouns and verbs close enough for them to be merged together.

3.2.1.2 Two adjectival classes

Whether Chinese has a separate adjective category has been debated for decades. Based on a line of reasoning similar to the above verb-as-noun analysis, some linguists argue that the so-called adjectives should be put under the verb category, since they can fill the predicator slot without any morphological marking (Li and Thompson 1989). Since verbs and most alleged adjectives show different morphological behaviors in reduplication, the verb-adjective distinction is kept, and the two are placed under the verbal category.

There still exist a (much smaller) number of alleged adjectives that shows different morphosyntactic properties with the adjectives in the verbal category (Paul 2014, chap. 5). They can be marginally used as heads of NPs, while they do not have reduplication variants. These “adjectives” are thus placed under the nominal category. Thus we have two types of adjectives. In 朱德熙 (2009), nominal adjectives are called 区别词 ‘distinction word’, while verbal adjectives are called 形容词 ‘adjective’.

3.2.1.3 Other nominal categories

There are more nominal categories than the ordinary noun category and the nominal adjective category. Numerals, for examples, are in another nominal category. Chinese has a rich classifier system, and most classifiers still have strong nominal properties and thus they constitute yet another nominal category. 朱德熙 (2009) calls them 量词 ‘measure word’, because many classifiers have the meaning of “unit”. There is also a locative particle class, including 里 in 在房子里, which is sometimes said to be the postposition class because they sometimes have adposition-like properties (TODO: ref: topicalization, and what else?).

3.2.2 Open and close classes; the lexical-functional distinction

3.2.2.1 The hierarchy of openness

The distinction between lexical and functional classes is sometimes subtle. (朱德熙 2009, § 3.6) classifies certain categories like locative particles into the nominal class and hence the lexical one, while the locative particle class can definitely be enumerated (朱德熙 2009, § 4.4). On the other hand, the author claims that lexical classes are always open and function classes are always closed 朱德熙 (2009, § 3.4). A conflict thus occurs.

The problem here is we have a gradient hierarchy from the prototypical lexical classes to the prototypical function classes. The most lexical class is open to new members, not a part of the grammar, and its members are able to be lexical heads (and thus has a “real part-of-speech label” like “noun” or “verb”) of, say, an NP or a verbal complex. A less open class is not so open to new members (just like Japanese verbs and adjectives), but is still not a part of the grammar and its members are able to be lexical heads. A even more closed class is not open to new members, and is a part of the grammar, but its members are still able to be lexical heads. Pronouns are in this type. A prototypical function class, then, is not open to new members, hardwired in the grammar, and its members are never lexical

heads. Derivational suffixes are in this type. This last type of forms bring no real part-of-speech label to its realization. We still classify these function items in the grammar into classes, but these classes are somehow less “real”.

It's of course not easy to tell a newly discovered part of speech (or *form class*, which may be a word class or an affix class) What's the status of an orientation preverb, which may be found in Japhug (Jacques 2021)? It's a part of the grammar, but does it carry a real part-of-speech label (like “directional adverb”)? And speaking of adverbs, what's the status of the English *allegedly*? An adverb filling a peripheral argument position, or an evidentiality marker? We really need to know a lot about language to fix the position of a form class. A common practice is just to shun the details and just say whether a class is lexical or functional, drawing a hard line between the two. So 朱德熙 (2009) mainly uses the criterion of whether there is a real part-of-speech label, and then directive particles are classified into the nominal class and they are in turn considered lexical. But he mistakenly confuses the notion of lexical classes with the notion of open classes, and then we get the self-conflicting asserts in 朱德熙 (2009, § 3.4).

3.2.2.2 Openness of Mandarin form classes

3.2.3 Summary: a tentative part of speech analysis

3.3 Nouns

3.3.1 Monosyllabic nouns

3.3.2 Nouns with historically analyzable inner structure

The unit 白菜 is made up by two perfectly productive morphemes: 白 ‘white’ and 菜 ‘vegetable’, but its meaning is not the composition of the two morphemes: 白菜 means ‘Chinese cabbage’, not ‘any vegetable with whitish appearance’. The word has already gained a conventionalized meaning, and its inner structure is of mostly diachronic interest but not synchronic interest. Therefore, the disyllabic unit 白菜 is the smallest unit fed into morphosyntax, and it of course is not a phrase.

Those insisting on the nonexistence of words in Chinese may explain the observation made above by claiming 白菜 to be an idiom NP: it is indeed a lexical entry, but is regarded as a pre-compiled phrase. TODO:

3.3.3 Nominal bound morphemes

3.3.4 Compound nouns

- (3) a. [[定义]_{modifier:N} [[等价]_{complement:adjective} [性]_{nominalizer}]_N]_N ‘equivalence of definitions’
- b. [[美国]_{modifier:N} [苹果]_{head:N}]_N
- c. *[美国]_{modifier:N} [红色的苹果]_{NP}
- d. 红色的美国苹果

From (3a) and (3b), it can be seen in certain morphosyntax units, a bare noun may serve as a (restrictive) modifier. The constituent of Chinese NPs is Dem Num A N, and this bare noun modifier position seems to be more internal than the adjective position, as is illustrated by (3c) and (3d). Furthermore, the bare noun position cannot be filled by an NP. The following examples demonstrate this:

- (4) a. [联合国] [秘书长]
- b. *[[某个组织] [秘书长]]
- c. 某个组织的秘书长

The obligatoriness of 的 means the NP 某个组织 can only appear as a modifier via the possessive construction. It cannot fill the slot of 美国 in 美国苹果. So the bare noun modifier position is a function label existing in a unit smaller than the NP – and it has to be the word.

3.4 Verbs

3.4.1 Verbs that are fossilized verb-object structures

Verbs, for example, may have inner predator-object structures. The verb 关心 ‘care for’ is certainly analyzable as a predator-object structure, but it takes objects just like any other verbs:

- (5) 他 [[[关]_{predicator:V} [心]_{object:N}]_{predicator:V} [自己的家人]_{object:NP}]_{predicate:VP}

This means 关心 is not a VP but a grammatical word, or otherwise it is impossible to take another object since there is no valency changing device in use. The verb 关心 itself poses no threat to the word-phrase distinction, but certain examples of these “words with internal syntax” can be tested to be words, while they indeed have phrasal counterparts. Compare the two examples below:

- (6) a. [念佛] 堂
b. 老太太 [念了这么久佛]_{VP1}, 却不知道自己在[念哪一尊佛]_{VP2}

The verb 念佛 in the first example is similar to 美国 in (3b): it serves as a bare modifier (since in Chinese verbal constituents can fill argument slots directly, the fact that 念佛 is a verb is not surprising). The fact 念佛 is able to appear in such a position assures us that it is a word. Then consider (6b). In VP1, a temporal semi-object is injected between the verb 念 and the object 佛, while in VP2, an interrogative phrase 哪一尊 is inserted into 佛 and an NP object is now taken by the verb 念. So here is the problem: what if 念佛 is *always* a VP, and what (6a) demonstrates is the bare noun (or verb) modifier may be sometimes filled by a phrase?

One solution – the solution taken by 朱德熙 (2009, § 1.2.6) – is to regard 念佛 in (6a) as a word, while the two VPs in (6b) as phrases. 念佛 as in 老太太经常念佛 can be interpreted as a word or as a phrase without making any difference. 念佛 as a word is something like *Buddha-praying*, while 念佛 as a verb is something like *pray to Buddha*. In the account of 冯胜利 (2000, p. 82), 念佛 is a morphosyntactic word (the original term being a 句法词 ‘syntactic word’), which is created by morphosyntactic rules and has an inner structure that is (partially) transparent for other morphosyntactic rules, while 关心 is a ‘lexical word’ (not the same with *lexical word* in the rest of this note which means words that are not a part of the grammar), which is taken out of the lexicon directly and has no synchronically analyzable inner structure. This also seems to be the analysis of 朱德熙 (2009), although less explicitly: he recognizes three classes of grammatical structures, 组合式 ‘composition-style’, 粘合式 ‘gluing-style’, and 复合词式 ‘compound word-style’, the first two being syntactic while the last class being morphological (§ 2.6.1) while recognizing that sometimes it’s hard to draw a line between the latter two (pp. 128-129). A gluing-up-style structure, in his definition, seems to be a regularly formed structure containing elements smaller than usual phrases – bare grammatical words? – and yet is a phrase itself, while a compound structure is within a grammatical word. So in this way a

3.5 Locational words

Monosyllabic location words like 前 may be analyzed as words because they can be attached to an arbitrary NP to denote a place near the place denoted by that NP, as in [[那座老旧的房子 [前]_{location word}]_{NP} 有一口井]_{clause}, and what only appears as immediate constituents of phrases are of course words,² but they never appear independently as NPs.

3.6 Prepositions

Though all Mandarin prepositions have verb origins and therefore may be classified as a subclass of verbs by some, it’s necessary to distinguish a separate preposition class. Criteria of prepositions include

TODO: ref

Box 3.3: The term *coverb*

In

²Here I have to emphasize again that I’m talking about morphosyntactic words here. It is possible that something is a morphosyntactic word is incorporated into a word nearby phonologically, though this is not the case in Chinese.

Chapter 4

The structure of noun phrase

No morphological case, number, and gender categories are attested in Mandarin. There is a word class system or in other words classifier system, however. In most cases when a numeral appears in an NP, a classifier follows immediately after the numeral. Attributives – both adjectives and relative clauses – follow the classifier. The demonstrative, if any, appears before the numeral, and even when there is no numeral, there is frequently also a classifier.

The template of NPs, therefore, belongs to the Dem-Num-A-N type, with the classifier residing between Num and A.

Chapter 5

The verbal complex

5.1 Introduction

Mandarin is generally regarded as a prototypical analytic language, without traditionally acknowledged verb inflections. Indeed it will be weird to posit something like a paradigm in Mandarin, but it doesn't mean there is no such thing as verbal affixation that are active in the morphosyntax (instead of not fully productive and arguably historical derivations). Some items involved here however may have partial mobility. Consider (1): In the first sentence, 了 is an aspectual suffix (§ 5.4), while 走 is a verb which never appear without an argument in uncontroversial phrasal grammar. So we conclude 了 and 走 are suffixes, and by structural comparison, we conclude 过来 in (1b) is also a suffix, with the same status as 走. But there comes (1c), in which 过来 moves to the end of the sentence.

- (1) a. 他 带 走 了 他的 文件
3sg carry go.away PERF 3sg-POSS file
'He carried his files away.'
- b. 他 带 [过来] 了 三 瓶 汽水
3sg carry come PERF three bottle.CL soda
'He carried here three bottles of soda.'
- c. 他带了三瓶汽水[过来]

To avoid the useless quarrelling about what is a word and whether a grammar point is morphology (which isn't that important in non-lexicalist generative theories, anyway), I use the term *verbal complex* to cover the main verb and the "suffixes" in (1). There are roughly three systems in the verbal complex. The first is the uncontroversial derivation system, like 化 '-ize'. The second is the verbal complement system, which includes three subsystems: the resultative complements, the directional complements, and the potential complements (§ 5.3). The third is the aspectual system (§ 5.4).

Box 5.1: On the notion of complements

The Chinese term 补语 corresponding to my *verbal complement* is frequently translated into the English term *complement*. This creates some confusion, because the term *complement* can also denote clausal dependents that are arguments of the main verb, as in Huddleston and Pullum (2002). The term *non-argument complement* may be used to avoid this confusion. There are, however, further confusions: Should we regard a clausal dependent that records the quantity or amount of an action as a non-argument complement? This construction can also be seen in Latin, like the Latin accusative expression of time (Greenough and Allen 2013, § 423). Thus, I use the term *verbal complement* to refer to things like 完 as in 做完了.

(2) is an example in which all the three systems appear. In real world speeches, such combinations have relatively lower distributions, possibly because of the prosodic constraint that verb shouldn't be too heavy unless it appears at the end of a clause (TDOO: ref).

- (2) ...并且企业 [数字 [化]_{derivation} [完]_{complement} [了]_{aspectual}]_V 之后还不一定赚钱 ...

Besides the systems shown in (2), the separation of a verb further complicates the behavior of the verbal complex (§ 5.5).

You may note the so-called serial verb constructions aren't mentioned here. Paul (2008) and 邓思颖 (2010, § 9.4) summarizes several constructions that are frequently referred to as serial verb constructions, and points out after deeper investigation, they can all be described in terms of the usual complement clause constructions, purpose clause constructions, etc. that are well attested cross-linguistically (§ 8.3.1).

5.2 Verbal derivations

5.3 Verbal complements

5.4 The aspectual system

5.4.1 TAME categories

Mandarin lacks the category of tense – all tense information is expressed by time adverbs. Modality is marked similarly by adverbs or complement clause constructions. Yet there is a system marking the aspect (§ 5.4). (3) is an example.

- (3) a. 我去过上海了
1 go GUO Shanghai SFP
'I have been in Shanghai.'
- b. 我去了上海了
1 go LE Shanghai SFP
'I have gone to Shanghai.'

标语贴在墙上 标语已经在墙上贴着了: this means the preposition 在 actually is morphologically merged with the verb 贴, or otherwise we are unable to explain why in the first example, 着 can never appear, while in the second example, 着 can appear.

5.5 Separable verbs

It's sometimes possible to split a verb and inject some clausal dependents into it. The interaction between this separation operation and the structure of the verbal complex is of some interest.

Chapter 6

Verb and arguments

Chapter 7

Valency changing

There are two ways of valency changing in Mandarin. The first is via a coverb construction, as in the disposal constructions (§ 7.1), TODO The second is *doing nothing* to the verb and relying on the unusual semantic roles of clausal complements to inform the listener about the valency changing, as in TODO: ref. Since there is no morphological marking, constructions of this type are often recognized as topic-comment structures, in which the “topic” – which is the subject under closer investigation – is said to be freely occupied by any semantic (and not necessarily syntactic) argument in the clause, though this claim can be falsified by detailed syntactic tests (§ 8.7.3).

7.1 The disposal constructions

7.2 The passive constructions

- (1) 我被他打了一拳

7.3 The causative

7.4 The affected construction

7.5 Instrumental object

- (2) 我们今天准备吃食堂

Chapter 8

Simple clauses

8.1 Overall remarks about the clause structure

A sentence can be divided into several clauses linked by clause linking constructions (chap. 11). This chapter is devoted to the simple clause, postponing details in subordination and clause linking to the next several chapters. Mandarin has rich topicalization 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. The comment – the nucleus clause – may further be divided into a subject (if any), a series of adverbials, the verbal complex, and post-verbal constituents, the most important types including object(s), the second part of a separable verb, certain directional complements, and purpose clauses.

Box 8.1: The term *clause*

Some people, like 邓思颖 (2010, p. 140) as well as Dixon (2009), use the term *clause* for subject-predicate constructions that don't receive complete marking of speech forces. (In generative terms, *clause* is for lower level CPs or even TPs.) So in this way, sentence-final particles shouldn't be discussed in this chapter because they are of course dependents in the sentence level. They may be discussed together with other sentence-level constructions like chap. 11. But this notion of clause certainly goes against the tradition in descriptive grammars. So the approach of this note is to acknowledge everything larger than TP as a clause, which may or may not be a sentence, and discuss its structure in this chapter, while “adjunctions” – or in other words, optional dependents – are discussed in, say, chap. 11, for the sake of convenience. The narrative order of this note is not the ideal “small unit – large unit” scheme, but the “simple large unit – complicated large unit” scheme. Needless to say, when it comes to clause combining, the problem of what the clause really is – with or without SFPs, for example – is still relevant, but it is not answered by saying “the construction takes a clause, not a sentence”.

As is implied by my using the term *subject*, Mandarin is an typical accusative language. Clausal dependents are recognizable from the rather rigid constituent order: Mandarin is usually classified as having a SVO clausal constituent order, and the subject and the object(s) can be told from the positions in the clause (1, 2). Certain “SOV” orders can be obtained by invoking the disposal construction (§ 7.1), as in (3).

- (1) 我生病 了
1 get.sick SFP
'I got sick.'
- (2) [我]_{subject} 今天 去 看 [电影]_{object} 了
1 today to watch movie SFP
'I went to watch a movie today.'
- (3) [我]_{subject} 今天 把 [一 个 碗]_{object} 摔 碎 了
1 today BA one CL bowl break crack SFP
'I broke one bowl today.'

The normal tests of syntactic accusative alignment can be run on Mandarin (4).

- (4) 陈 经理 昨天 没有和 他的 客户 聊 过 。 他 生病 了 。
 Chen (surname) manager yesterday NEG with 3sg-POSS client talk SFP 3sg get.sick SFP
 ‘Manager Chen didn’t talk with his client yesterday. He (Chen, not his client) got sick.’

8.2 Prosodic constraint

The main verb and one post-verbal constituent should form the last prosodic constituent in the clause (ignoring SFPs), and when there is no post-verbal constituents, the main verb receives the natural stress. This is actually a rather strong condition. Certain constituents – most functional words – are unable to accept stress at all. They may be freely merged into the closest prosodic constituents. Certain constituents – like NPs with definite references – are by default stressed. When the verb is the last lexical constituent (as in *他挥动棒子把我打), it should not be too short or otherwise it is unable to receive stress. When there are more than one post-verbal constituents, only one of them can receive stress. If two post-verbal constituents are both by default stressed, the sentence is again ruled out.

8.3 Types of nucleus clauses

Box 8.2: About the subject-predicate binary division

Dixon (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 in the topic-comment construction.

8.3.1 There is no serial-verb construction or complex predicate

8.4 Splitting of the verb

Chao (1965, § 6.5.8) records the first two non-standard examples of the phenomenon in the follows, while the third example is more widely accepted:

- (5) a. % [军完了训] 以后才可以去请护照
 b. % 还 [幽了他一默]
 c. 这件事情你 [关什么心] 啊

The bracketed constituents in (5b) and (5c) are both uncontroversially VPs: (5b) contains the object 他 and a temporal semi-object 一, while in (5c), 什么心 as an NP is the object. The splitting of the verbs clearly origins by analogy with VPs containing morphosyntactic words. The bracketed constituent in (5a) is equivalent to 军训完了, the latter being a verb complex, but 军完了训 is apparently created by analogy with [[吃完了]_{predicator:verb complex} [饭]_{object:N}]_{VP}. Anyway, whether 军完了训 is directly a VP or is first a verb complex and then a VP is not of much importance: similar construction never when an object is present. What can be seen from (5), then, is uncontroversial grammatical words can also be split and extended into phrases in the same way phrase-like words like 念佛 do in § 3.1.3.

Splitting a word is of course morphological, and is of a nontrivial type, but here in Chinese, what is injected in after splitting the verb are phrasal dependents and the resulting morphosyntactic unit is a phrase, so what we find here is syntax seems to rely on morphological devices. Without prejudice, this is not quite shocking, since what are said to be head movements also involve fusion in Distributed Morphology and therefore may be described as morphological.

The motivation of this phenomenon seems to be prosody: splitting words into phrases is only observed in VPs, and VPs are subject to the prosodic constraint that the neither the verb nor the final complement can be too light. Splitting the verb may help to reduce the “weight” of the verb so the resulting utterance meets the prosodic constraint better.

It should be noted that some verb ionization examples seem to be analyzable as trivially extending the object: compare 看书 and 看了三本书. Some verbs with verb-object inner structures contain an internal object that usually doesn’t appear as a full NP, but that is not absolute: 洗了一个舒服的澡

and 一个舒服的澡对睡眠有好处. Prototypical verb ionization seems to be restricted to semi-objects (TODO: ref) 理了三次发.

8.5 Negation

Like the case in standard English, there is no negative concord in Mandarin Chinese. There is, however, no uniform negation operator like the English *not*. Several negation operators and strategies are used frequently (§ 8.5). Verbs can be negated by 不 while nouns generally cannot, and this is a criterion to tell verbs from nouns. There is another negation operator 没, which has subtle differences in its meaning and syntactic properties compared with 不 (6, 7). On the other hand, the negative potential complement construction, i.e. the V不了 construction, isn't obtained by inserting a negator in the clause (8).

- (6) a. 我 不 喜欢 吃 芹菜
1 NEG like eat celery
'I don't like eating celery.'
- b. * 我没喜欢吃芹菜
- (7) a. 我 不 吃 早饭
1 NEG eat breakfast
'I don't eat breakfast. (I usually don't, I don't want any today, etc.)'
- b. 我 没 吃 早饭
1 NEG eat breakfast
'I didn't eat breakfast. (I may usually do, but somehow I didn't today.)'
- (8) a. 我做 [不 了]_{potential complement, negative} 这 件事
1 do NEG finish this CL affair
'I'm not able to do this.'
- b. * 我 没有/并非/不 做 [得 了]_{potential complement, positive} 这 件事
1 NEG do DE finish this CL affair

8.6 Sentence final particles

8.7 The topic-comment structure

I follow Shi (2000)'s approach and define a topic as an unmarked NP that has certain relations with a position in the clause after it and is indeed the topic in the information structure (i.e. some (probably already known) object to which new information is added). Constructions like 连...都...are not discussed in this section – they are to be found in TODO: ref.

8.7.1 Topicalization of possessor

(9a) and (9a) are a pair of sentences with and without topicalization of the possessor in the subject.

- (9) a. [他]_{topic} [[个子]_{subject} 高高 的]_{comment}
3sg stature tall~TODO DE
'As for him, the stature is tall.'
- b. [他 的 个子]_{subject} 高高 的
3sg POSS stature tall~TODO DE
'His stature is tall.'

8.7.2 Topicalization of preposition objects

- (10) 这件事你不能就麻烦他一个人
- (11) 你不能[为了这件事]_{adverbial:PP} 就麻烦他一个人

This is also a demonstration of the preposition status of 在 in this sentence (§ 3.6), because if it's a verb or an auxiliary verb, it will be hard to have its object topicalized and have it deleted at the same time, but deletion of the preposition in topicalization is well-attested cross-linguistically.

8.7.3 Rejecting the notion of dangling topics

Some people, like 朱德熙 (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 is just an NP which the comment is “about”, 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.

8.7.3.1 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. Such cases are however rather unproductive. In (12) and (13), the orders of the constituents can never be changed. Nor is it possible to change a word or two in the bracketed “comments”. A reasonable assumption is these bracketed “comments” are actually idioms, which are to be regarded as a single verbal element that can't be further analyzed. Thus, in (12) and (13), the so-called topic is an ordinary subject, and the so-called comment is a predicate.

(12) 他们[大鱼吃小鱼] (， 厮杀成一片)

(13) 他们[你看看我我看看你]

8.7.3.2 Type 2: Quantificational adverbial looking like the inner subject

The second type of “dangling topic” is like (14). A topic-comment analysis of (14)

(14) 他们 谁 都 不 怕
3pl who even NEG fear
'They don't fear anyone.'

8.7.3.3 Type 3: Ellipsis leaving a subject and one predicate

Some people accept (15). 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 (15): (16) is a demonstration of the 幸亏……不然…… linking construction, and we also have its topicalized version (17). (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 (15) is clear: We can get it by omitting the second clause in the comment part of (17). 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.

(15) % 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).'

(16) [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.'

(17) [那 所 房子]_{topic} [幸亏 去年 没 下雪 , 不然 早就 塌 了]_{comment}
that CL house fortunate last.year NEG snow otherwise already collapse SFP

8.7.3.4 Type 4: Extraction from prepositional adverbials

(10) in § 8.7.2 is sometimes regarded as an instance of the dangling topic construction. However, as is shown in § 8.7.2, it may just be from topicalization of an NP in an adverbial, with the preposition (and/or the locative particle) removed.

8.7.3.5 Type 5: Nominal predicate

- (18) 这种青菜一斤三十块钱

8.7.3.6 Type 6: Locational adverbial mistaken for the subject

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

8.7.3.7 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 structure. 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 8.3: 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

Chapter 9

Relative clause constructions

Chapter 10

Complement clause constructions

Box 10.1: Non-existence of finite-nonfinite distinction in Mandarin

Cross-linguistically, we find a finite-nonfinite distinction in subordination. This distinction is arguably absent in Mandarin, even after detailed syntactic tests (Hu et al. 2001).

Chapter 11

Clause linking

Mandarin Chinese has usual clause linker devices (chap. 11), as well as complement clauses (§ 10) and relative clauses (TODO: ref). TODO: what else?

References

- Yuen Ren Chao. *A grammar of spoken Chinese*. ERIC, 1965.
- Robert MW Dixon. *Basic linguistic theory volume 1: Methodology*, volume 1. OUP Oxford, 2009.
- James B Greenough and Joseph H Allen. *Allen and Greenough's new Latin grammar*. Courier Corporation, 2013.
- Jianhua Hu, Hai hua Pan, and Liejiong Xu. Is there a finite-non-finite distinction in chinese? *Linguistics*, 39:1117–1148, 01 2001. doi: 10.1515/ling.2001.043.
- Rodney Huddleston and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge University Press, 2002. doi: 10.1017/9781316423530.
- Guillaume Jacques. *A grammar of Japhug*, volume 1. Language Science Press, 2021.
- Charles N Li and Sandra A Thompson. *Mandarin Chinese: A functional reference grammar*, volume 3. Univ of California Press, 1989.
- Waltraud Paul. The serial verb construction in chinese: A tenacious myth and a gordian knot. *The Linguistic Review*, 2008.
- Waltraud Paul. *New perspectives on Chinese syntax*. De Gruyter Mouton, 2014.
- Dingxu Shi. Topic and topic-comment constructions in mandarin chinese. *Language*, 76, 06 2000. doi: 10.2307/417661.
- Shiao Wei Tham. Resultative verb compounds in mandarin. *The Oxford handbook of Chinese linguistics*, pages 306–321, 2015.
- Andreas Trotzke. Constructions in minimalism: A functional perspective on cyclicity. *Frontiers in Psychology*, 11, 2020. ISSN 1664-1078. doi: 10.3389/fpsyg.2020.02152. URL <https://www.frontiersin.org/article/10.3389/fpsyg.2020.02152>.
- 冯胜利. 汉语韵律句法学. 汉语韵律句法学, 2000.
- 朱德熙. 语法讲义. In 朱德熙文集, 第一卷. 商务印书馆, 2009.
- 王理嘉, 陆俭明, 符淮青, 马真, and 苏培成. 现代汉语(重排本). 商务印书馆, 2004.
- 邓思颖. 形式汉语句法学. 形式汉语句法学, 2010.