The Cambridge Handbook of

Chinese Linguistics

edited by **Chu-Ren Huang,
Yen-Hwei Lin, I-Hsuan Chen
and Yu-Yin Hsu**

The Cambridge Handbook of Chinese Linguistics

Edited by

Chu-Ren Huang
The Hong Kong Polytechnic University

Yen-Hwei Lin
Michigan State University

I-Hsuan Chen
University of California, Berkeley

Yu-Yin Hsu
The Hong Kong Polytechnic University

CAMBRIDGE
UNIVERSITY PRESS

Contents

Figures

Tables

5

Parts of Speech in Chinese and How to Identify Them

Weidong Zhan and Xiaojing Bai

5.1 Introduction

In constituent-based language models, words[1] combine to form phrases and sentences. What are the principles or rules governing the process of combination? To better answer the question, we need to distinguish words based on their grammatical functions, place them into different classes, or parts of speech, and describe the constraints of their combination clearly by referring to their parts of speech.

Language is a system of meaning–form pairs. The role of a word is determined by its meaning and the syntactic properties derived from its meaning (Shi 石定栩 2009; Huang and Shi 2016). Preferably, words in one class function similarly and thus share semantic and syntactic properties, while words from different classes do not.[2] Such is the case with words in (1), where four parts of speech are involved: numeral (expressing a number, e.g., 一 *yi* 'one', 两 *liang* 'two', 三 *san* 'three'), classifier (indicating the type, shape, or other properties associated with a noun, e.g., 辆 *liang*, 所 *suo*, 棵 *ke*), noun (identifying a class or a particular one of people, places, or things, e.g., 汽车 *qiche* 'motor vehicle', 大学 *daxue* 'university', 柳树 *liushu* 'willow'), and demonstrative (alluding to a near or far reference, e.g., 这 *zhe* 'this', 那 *na* 'that', 其他 *qita* 'other'). Examples in (1a) and (1c) are all grammatical noun phrases, as Numeral-Classifier-Noun constructions in (1a) and as Demonstrative-Numeral-Classifier-Noun constructions in (1c). Examples in (1b) and (1d) are ungrammatical sequences of words (in the patterns of

---

[1] The wordhood issue is complex in Chinese, largely due to the lack of word delimiters (i.e., spaces between words). In this chapter, a word refers to the smallest unit that has a meaning and fills a particular position in a sentence (Chao 1968).

[2] This is the preferred or ideal case for Chinese and all other languages, which is what general linguistics or universal grammar seeks to formulate. If words do behave that way, the typological comparisons of parts of speech across languages would become straightforward, and the learning of a foreign language would bear more resemblance to the acquisition of one's mother tongue.

Numeral-Noun and Numeral-Classifier-Demonstrative-Noun, respectively). Using parts of speech, a small number of rules can be written to describe a large (or even an infinite) number of legitimate combinations of words.

(1)   a. 一 辆 汽车   两 所 大学   三 棵 柳树
          yi__liang__qiche   *liang__suo__daxue*   *san__ke__liushu*
          one car   two universities   three willows
      b. * 一 汽车   * 两 大学   * 三 柳树
          * *yi__qiche*   * *liang__daxue*   * *san__liushu*
      c. 这 一 辆 汽车   某 两 所 大学   那 三 棵 柳树
          zhe__yi__liang__qiche   *mou__liang__suo__daxue* *na__san__ke__liushu*
          this car   two universities   those three willows
      d. * 一 辆 这 汽车   * 两 所 某 大学   * 三 棵 那 柳树
          * *yi__liang__zhe__qiche* * *liang__suo_mou_daxue* * *san__ke__na__liushu*

This chapter focuses on the parts of speech in Chinese. Section 5.2 introduces the principles of defining parts of speech in Chinese and Section 5.3 gives the details of how they are defined. The overlapping of parts of speech is examined in Section 5.4, with an extended discussion about handling the overlapping classes. Section 5.5 concludes the chapter and references other studies on the same issue but from different perspectives.

5.2   Principles of Defining Parts of Speech in Chinese

While there is not much argument about classifying words in Chinese based on their semantic and syntactic properties, there is no consensus on what properties in particular are more distinguishing and are therefore to be used as the definitional criteria for parts of speech. Consequently, different classification systems have been proposed, with diverging criteria and varying parts of speech.

For instance, nouns are words used to identify things, animals, plants, institutions, people, etc. or to name a particular one of them. However, it is not settled whether words denoting time and location are to be included, which leads to different definitions of nouns. Syntactically, a part of speech can be defined by the functions, i.e., positions, of the words in a sentence, as is the traditional way in English and many other Indo-European languages. Otherwise, the definitional criteria can be the functions of the words in phrase formation, which applies well to the Chinese language.

Concerning the parts of speech in Chinese and the principles of defining them, some common positions have been reached so far among researchers.

First, both semantic and syntactic properties of words are necessarily considered, highlighting the similarity among words of the same class as much as possible and distinguishing words of different classes to the largest extent. With content words, more consideration is given to the semantic

Table 5.1. *Examples of typical phrasal constructions in Chinese*

Phrasal construction	Examples
Subject-Predicate	老师_同意 *laoshi_tongyi* 'the teacher agrees'
	价格_高 *jiage_gao* 'the price is high'
Verb-Object	学习_语法 *xuexi_yufa* 'to learn the grammar'
	写_论文 *xie_lunwen* 'to write a paper'
Verb-Complement	打听_清楚 *dating_qingchu* 'to ask and confirm'
	洗_干净 *xi_ganjing* 'to wash clean'
Adverbial-Head	赶快_出发 *gankuai_chufa* 'to set out immediately'
	都_错 *dou_cuo* 'to be wrong completely'
Attributive-Head	红_裤子 *hong_kuzi* 'red trousers'
	大型_企业 *daxing_qiye* 'big enterprise'
Numeral-Classifier-Head	一_辆_汽车 *yi_liang_qiche* 'one car'
(Attributive-Head)	两_所_大学 *liang_suo_daxue* 'two universities'
很-Head[a]	很_高 *hen_gao* 'very high'
(Adverbial-Head)	非常_喜欢 *feichang_xihuan* 'to like very much'
Noun-Locative	桌子_旁边 *zhuozi_pangbian* 'by the table'
	教室_里 *jiaoshi_li* 'in the classroom'

[a] In this construction, 很 *hen* 'very' is representative of the adverbs that are used similarly to express high degree, such as 非常 *feichang* 'very much', 特别 *tebie* 'particularly', and 挺 *ting* 'quite'.

properties; with function words, more emphasis is placed on the syntactic ones.

Second, in addition to sentences, phrasal constructions are usually used to describe the syntactic properties of words in Chinese. A sentence can be divided into its immediate constituents, i.e., phrasal constructions, which in turn consist of their own immediate constituents. Therefore, it is appropriate to define a part of speech syntactically by describing its positions in different phrasal constructions. (See Table 5.1 for more details of the typical phrasal constructions in Chinese.)

Third, the semantic and syntactic criteria can be coarse-grained or fine-grained, which will cause relativity in the definitions of parts of speech. For instance, time and location are concepts closer in meaning to thing than to action and attribute. Therefore, words denoting time and location, such as 明天 *mingtian* 'tomorrow', 元旦 *yuandan* 'New Year's Day', and 郊区 *jiaoqu* 'suburb', can be defined as nouns, rather than verbs or adjectives. A closer look at the meaning of these words, however, sets those denoting time and location apart from others denoting entities, such as 书 *shu* 'book', 狗 *gou* 'dog', 花 *hua* 'flower', 学校 *xuexiao* 'school', 教授 *jiaoshou* 'professor', etc. Thus, time words and location words are defined as parts of speech separate from nouns.

Fourth, parts of speech are usually not defined all at one time. Some are easily and hence initially confirmed, which may help to define others. In general, parts of speech with simpler and clearer semantic or syntactic properties are first defined, which tend to be closed classes with a small number of members, such as interjections, conjunctions, prepositions, etc.

Other parts of speech are defined one after another, with increasing complexity of the definitional criteria in most cases. They tend to be large in number, including many open classes that continually accept new members, such as nouns. However, flexible decisions are often made when parts of speech are defined. In Lu 陆俭明 (1994, 2005), for instance, interjections are grouped together at the very beginning, as they function independently and are small in number, such as 啊 *a* 'ah', 哎 *ai* 'well', 嗯 *en* 'eh', 嚯 *huo* 'wow', 哦 *o* 'oh', and 呀 *ya* 'oh'. Conjunctions are defined next, which are small in number as well but function more diversely. When used for connection, conjunctions may occur in different positions: between two elements, such as 和 *he* 'and' and 并且 *bingqie* 'and'; before an element, such as 如果 *ruguo* 'if' and 虽然 *suiran* 'although'; and after an element, such as 的话 *dehua* 'if'. Following conjunctions, other parts of speech are defined one by one, among which classifiers, classifying words, and adverbs are defined at last. Yet these are three relatively small and closed classes, and they could have been defined in earlier steps otherwise.

5.3   Definition of Parts of Speech in Chinese

5.3.1   Typical Phrasal Constructions in Chinese

As has been discussed previously, the meaning of a word contributes to its grammatical function and can therefore be used to define its part of speech. However, meaning is primarily perceived intuitively, which involves subjectivity. Without formalized and verifiable means, semantic criteria may give rise to ambiguity, inconsistency, or other problems if used to define parts of speech. Such being the case, phrasal constructions are formalized so as to describe the positions that a class of words can or cannot take, and the definitional criteria for parts of speech in Chinese can thereby be established. Table 5.1 lists some typical phrasal constructions in Chinese, which will be referred to in Table 5.2 when parts of speech are defined.

In the grammar of Chinese, syntactic knowledge can in the first place be represented by such phrasal constructions. There are typical properties of each construction and distinctions between different ones, which are further discussed by Zhu 朱德熙 (1982) and Zhan 詹卫东 (2000), namely, what kinds of meaning a phrasal construction is to express and how one construction can be transformed into another.

Each phrasal construction has selectional restrictions on its constituents, which gives us a way to define the parts of speech in Chinese. For example, adjectives can occur as the head in the construction 很-Head, while nouns, verbs, stative words, and classifying words cannot. Table 5.2 shows that most parts of speech can be clearly defined by their positions in the constructions listed in Table 5.1.

However, it is worth noting that the definitional criteria in Table 5.2 do not necessarily guarantee the internal consistency and the mutual

exclusiveness of the resulting parts of speech, which will be further discussed in Section 5.4. Essentially, a part of speech can be viewed as a subset of words bearing family resemblance and represented by prototypical members (Yuan 袁毓林 2010). It is reasonable to define a certain part of speech by the positions that its prototypical members can take in phrasal constructions, but it is not necessary to require each word of this class to behave exactly the same way.

Further, although most of the parts of speech in Table 5.2 are defined according to their positions, some are otherwise defined, of which function words are typical examples. Obviously, there is no position that a function word can take in the typical constructions listed above.

Ever since the publication of the first grammar of Chinese, *Ma's Grammar*, in 1898, there has been a mountain of scholarship on the classification of Chinese words in the field of Chinese grammar research. Various classifications have been proposed (Lu 陆俭明 2005; Xu and Tan 徐枢, 谭景春 2006), with a growing interest in a finer granularity of classes[3] and a deeper understanding of their syntactic properties. Meanwhile the number of shared parts of speech is also growing across different classifications, which suggests an increasing consensus among researchers.[4] Of the existing classifications, the one to be introduced below includes the largest number of classes and presents many shared and in-depth understandings of parts of speech in Chinese. For more discussion about the disagreement on word classification, see Zhan 詹卫东 (2013).

5.3.2 A Hierarchy of Word Classification in Chinese

The classification of Chinese words, proposed by the Department of Chinese Language and Literature of Peking University (Zhu 朱德熙 1982; Guo 郭锐 2002; Guo 郭锐 2014), makes the finest distinctions between the parts of speech in Chinese[5]. The classification is represented as a horizontal hierarchical tree structure in Figure 5.1, which displays the distinctions between words on three levels.

First, words that cannot be used in combination with other words are distinguished from those that can. Interjections are the most typical examples, which always form sentences independently.

Second, words that combine to form a sentence are divided into content words and function words. The former can take at least one position in a

---

[3] Following the Western research paradigm, *Ma's Grammar* defined nine classes of words. The number increased substantially a hundred years later, as 20 classes were proposed by Guo (2002) and 17 adopted in the *Contemporary Chinese Dictionary*.

[4] With the eight classification systems practically used in the past two decades, 14 parts of speech are generally acknowledged, which contrasts with the situation in the last century, when 11 systems prevailed, with only eight parts of speech commonly accepted.

[5] In natural language processing, PoS tags tend to be more fine-grained, which may also involve punctuations, semantic labels, discourse markers, etc.

typical phrasal construction and have concrete lexical meaning. The latter do not occur in any typical phrasal construction but do perform grammatical functions, and their lexical meaning is hard to define. The former are open classes, with new members continually coined or borrowed from other languages. The latter are closed classes, with relatively fixed and thus enumerable members. As a result, there are more content words than function words.

Third, content words are further divided into nominals and predicatives.[6] The former are used to refer to entities of all kinds, and the latter used to predicate or describe the entities as performing a certain action or as having a certain property.

5.3.3   Definitional Criteria for the Parts of Speech in Chinese

The leaf nodes of the tree structure in Figure 5.1 are the 20 parts of speech in Chinese, which can be identified by the semantic and syntactic properties specified in Table 5.2. For a class of content words, semantic properties refer to their lexical meaning in general; and for a class of functions words, their grammatical meaning. What functions a class of words can or cannot play in the typical constructions (see Table 5.1) are viewed as their syntactic properties, of which only the most distinctive ones are listed.

Some issues are worth pointing out concerning the definitional criteria listed above.

First, the identification of parts of speech can be quite intuitive for humans. This is particularly true with the words that have clear lexical meanings. In such cases, syntactic criteria are simply additional means of verification.

Second, syntactic properties as those in Table 5.2 describe the functions of words in a finer way and can thus help to distinguish words with similar lexical meanings.

(2)   a.   他**迅速**销毁了密码本
             ta__xunsu__xiaohui__le__mima__ben
             He quickly destroyed the password book.
       b.   他销毁密码本很**迅速**
             ta__xiaohui__mima__ben__hen__xunsu
             He destroyed the password book very quickly.

(3)   a.   他**立即**销毁了密码本
             ta__liji__xiaohui__le__mima__ben
             He immediately destroyed the password book.
       b.   * 他销毁密码本很**立即**
             * *ta__xiaohui__mima__ben__hen__liji*

---

[6] In Figure 5.1, the dotted lines signify that pronouns and onomatopoeias can function as nominals and predicatives in different contexts. In addition, Guo 郭锐 (2002, 2014) distinguish three types of content words, namely nominals, predicatives, and modifiers.

Table 5.2. *Definitional criteria for the parts of speech in Chinese*

Part of speech	Semantic properties	Syntactic properties	Examples
Numerals	Expressing an amount or used for counting and numbering	Functioning as Numeral in Numeral-Classifier-Head	一 *yi* 'one' 两千 *liangqian* 'two thousand' 多 *duo* 'many' 第十 *dishi* 'tenth'
Classifiers	Expressing a unit of an amount or of measuring and counting	Functioning as Classifier in Numeral-Classifier-Head	个 *ge* 张 *zhang* 只 *zhi* 套 *tao*
Quantity words	Expressing an approximate amount or the whole of a particular group or thing without mentioning the unit	Functioning as Numeral + Classifier[7] in Numeral-Classifier-Head	大量 *daliang* 'a large amount of' 多数 *duoshu* 'majority' 一切 *yiqie* 'all' 所有 *suoyou* 'all'
Nouns	Expressing a concrete or abstract entity	Functioning as Head in Numeral-Classifier-Head	电脑 *diannao* 'computer' 作家 *zuojia* 'writer' 马 *ma* 'horse' 语言 *yuyan* 'language'
Classifying words	Expressing an attribute of an entity for the purpose of classification	Functioning as Attributive in Attributive-Head; not as Head in Numeral-Classifier-Head or 很-Head; not as Predicate in Subject-Predicate; not as Verb or Complement in Verb-Object or Verb-Complement	男 *nan* 'male' 女 *nv* 'female' 大型 *daxing* 'large-scale' 主要 *zhuyao* 'primary'
Locative particles	Expressing a relative direction and location	Functioning as Locative in Noun-Locative	上 *shang* 'higher place' 前 *qian* 'front' 右 *you* 'right' 旁边 *pangbian* 'side'
Location words	Expressing a location	Functioning not as Noun in Noun-Locative	民间 *minjian* 'among people' 当地 *dangdi* 'in the locality' 远处 *yuanchu* 'in the distance' 四周 *sizhou* 'all around'
Time words	Expressing time	Functioning as Noun in Noun-Locative[8]	春节 *chunjie* 'the Spring Festival' 今年 *jinnian* 'this year' 现在 *xianzai* 'now' 周末 *zhoumo* 'weekend'

[7] Quantity words modify nouns without taking any classifier, which sets them apart from numerals, and are therefore marked as Numeral + Classifier here.

[8] The Noun-Locative construction is given here to define time words, as they are found in similar usages as location words, for example, 春节_前 *chunjie_qian* 'before the Spring Festival' and 春节_后 *chunjie_hou* 'after the Spring Festival', which are formed in the same way as 矮墙_前 *aiqiang_qian* 'in front of the low wall' and 矮墙_后 *aiqiang_hou* 'behind the low wall'. The locative particles 前 *qian* and 后 *hou* help to indicate both the sequence in time and the location in space.

Table 5.2. (*cont.*)

Part of speech	Semantic properties	Syntactic properties	Examples
Demonstratives	Indicating the entity referred to	Functioning as Attributive in Attributive-Head, where Head is a Numeral-Classifier-Head construction[9]	任何 *renhe* 'any' 某 *mou* 'a certain' 这 *zhe* 'this' 其他 *qita* 'other'
Pronouns	Referring to what has been previously mentioned	Functioning as the expression previously used for the same referent[10]	我 *wo* 'I; me' 这里 *zheli* 'here' 哪儿 *naer* 'where' 这么 *zheme* 'in such a way'
Onomatopoeia	Imitating the sound	Functioning as Adverbial in Adverbial-Head; as Attributive in Attributive-Head	嗡嗡 *wengweng* 'hum' 叽里咕噜 *jiligulu* 'babble' 砰 *ping* 'ping' 滴答 *dida* 'tick'
Verbs	Expressing an action, behavior, relation, etc.	Functioning as Predicate in Subject-Predicate; as Verb in Verb-Object or Verb-Complement; not as Head in 很-Head	吃 *chi* 'to eat' 支持 *zhichi* 'to support' 属于 *shuyu* 'to belong to' 有 *you* 'to have'
Adjectives	Expressing an attribute	Functioning as Predicate in Subject-Predicate; as Complement in Verb-Complement; as Head in 很-Head; not as Verb in Verb-Object	白 *bai* 'white' 干净 *ganjing* 'clean' 认真 *renzhen* 'serious' 紧张 *jinzhang* 'nervous'
Stative words	Expressing a state	Functioning as Predicate in Subject-Predicate; as Complement in Verb-Complement; not as Head in 很-Head; not as Verb in Verb-Object	雪白 *xuebai* 'snow-white' 干干净净 *ganganjingjing* 'spotless' 香喷喷 *xiangpenpen* 'yummy' 巨大 *juda* 'giant'
Adverbs	Used to modify a verb or an adjective	Functioning as Adverbial in Adverbial-Head	居然 *juran* 'unexpectedly' 曾经 *cengjing* 'once' 很 *hen* 'very' 马上 *mashang* 'immediately'
Prepositions	Introducing an entity involved in an action	Functioning as Adverbial in Adverbial-Head when taking words that express the concepts of entity, event, time, location, etc.	把 *ba* 被 *bei* 'by' 从 *cong* 'from' 对 *dui* 'to'

[9] This criterion sets demonstratives apart from adjectives or quantity words, because an adjective or quantity word cannot modify a Numeral-Classifier-Head construction. It is appropriate to say 这_两_个_人 *zhe_liang_ge_ren* 'these two persons', but not *优秀_两_个_运动员 *youxiu_liang_ge_yundongyuan*, nor *很多_两_个_家庭 *henduo_liang_ge_jiating*. 任何 *renhe* 'any' and 某 *mou* 'a certain', with indefinite connotations though, are placed in this class, as they behave like other demonstratives, with the same syntactic properties.

[10] Pronouns in Chinese can be nominal or predicative. A nominal pronoun replaces an expression in the context to refer to the same entity, number, place, time, etc., such as 这里 *zheli* 'here'; a predicative pronoun replaces an expression in the context to refer to the same action, behavior, manner, etc., such as 这么 *zheme* 'in such a way'. Thus, the syntactic properties that can define pronouns are not their own distributions, but the distributions of the expressions they replace. In natural language processing, pronouns are usually subcategorized, so that there may also be pro-numerals (e.g., 几 *ji* 'a few; how many', 多少 *duoshao* 'how much; how many'), pro-adverbs (e.g., 这么 *zheme* 'in such a way', 这样 *zheyang* 'in this way'), pro-verbs (e.g., 怎么样 *zenmeyang* 'do as one likes'), etc.

Table 5.2. (*cont.*)

Part of speech	Semantic properties	Syntactic properties	Examples
Conjunctions	Connecting constituents and marking the logical relation between them	Taking a variety of positions and connecting constructions	和 *he* 'and' 不但 *budan* 'not only' 否则 *fouze* 'or' 以免 *yimian* 'lest'
Sentence-final particles	Expressing an attitude or indicating the illocutionary force	Taking the sentence-final position	吧 *ba* 吗 *ma* 呢 *ne* 呀 *ya*
Other particles	Expressing a variety of grammatical meanings	Taking a variety of positions and attached to constructions	的 *de* 得 *de* 了 *le* 着 *zhe*
Interjections	Expressing an emotion or used as a response when being called	Functioning as an independent sentence	哎呀 *aiya* 'wow' 哎哟 *aiyo* 'ouch' 喂 *wei* 'hello' 嗨 *hai* 'hi'
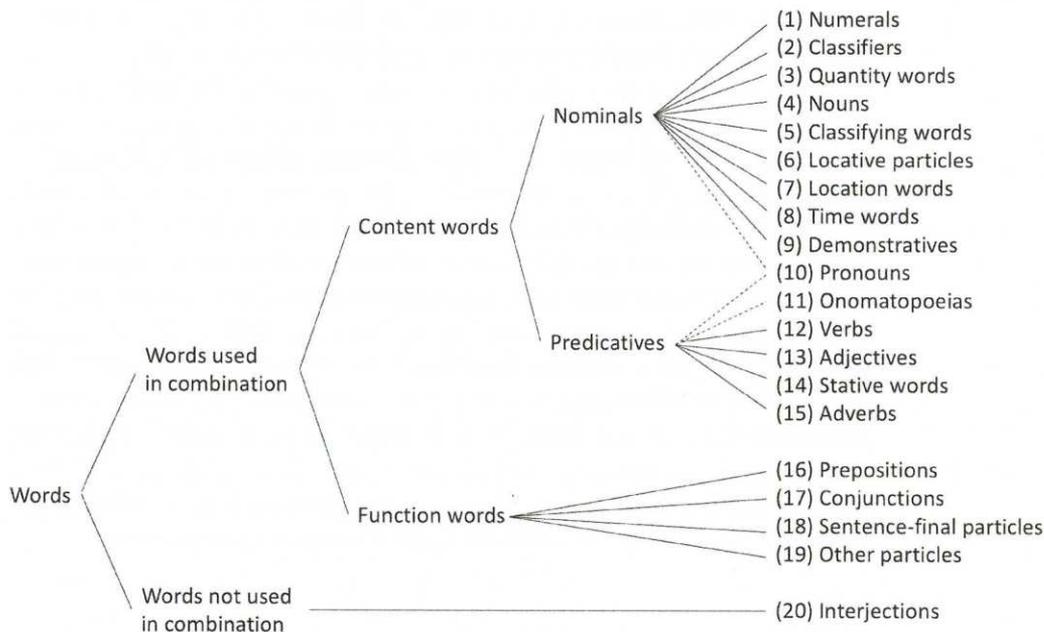
Figure 5.1 A hierarchy of word classification in Chinese

In (2a) and (3a), for example, the two words 迅速 *xunsu* 'quick' and 立即 *liji* 'immediately' are close in meaning, namely '(do sth.) with little or no delay', and their grammatical functions are exactly the same. However, an obvious distinction can be found between them in (2b) and (3b), as only 迅速 *xunsu* can combine with 很 *hen* 'very' to form the construction 很-Head. Thus, 迅速 *xunsu* is identified as an adjective, and 立即 *liji* an adverb.

Third, only a brief and incomplete account of syntactic properties is provided for each part of speech in Table 5.2. For instance, not all the positions that onomatopoeias can take are given, and with classifying words and location words, the descriptions are mainly about the positions that they cannot take. Further, many syntactic properties are described with an aim of highlighting the difference between parts of speech that have similar properties in meaning, so that classifying words can be set apart from nouns and adjectives, location words distinguished from nouns and locative particles, and stative words separated from adjectives and verbs. For more details about the distributions of parts of speech in Chinese, see Guo 郭锐 (2002, 2014) and Yu et al. 俞士汶等 (2003).

Fourth, the account of syntactic properties in Table 5.2 focuses on the prototypical members in each class. For instance, nouns are defined as the words that can be the Head in the construction Numeral-Classifier-Head, such as the words 电脑 *diannao* 'computer', 作家 *zuojia* 'writer', and 马 *ma* 'horse' in 一_台_电脑 *yi_tai_diannao* 'a computer', 两位作家 *liang_wei_zuojia* 'two writers', and 三匹马 *san_pi_ma* 'three horses'. However, there are still nouns that do not meet this definitional criterion, such as 年龄 *nianliang* 'age', 笔者 *bizhe* 'author', 对方 *duifang* 'the other party', and 苍天 *cangtian* 'Heaven', which is quite common in the classification of words.

Finally, some parts of speech can be subcategorized. For instance, there are transitive verbs (e.g., 吃 *chi* 'to eat', 喝 *he* 'to drink'), intransitive verbs (e.g., 咳嗽 *kesou* 'to cough', 休息 *xiuxi* 'to rest'), directional verbs (e.g., 上 *shang* 'to go up', 下 *xia* 'to go down', 来 *lai* 'to come', 去 *qu* 'to go'), verbs taking a nominal object (e.g., 敲 *qiao* 'to knock', 踢 *ti* 'to kick'), verbs taking a predicative object (e.g., 希望 *xiwang* 'to hope', 反对 *fandui* 'to oppose'), etc. There are individual classifiers, aggregate classifiers (e.g., 对 *dui* 'pair', 双 *shuang* 'pair', 伙 *huo* 'group'), unit classifiers (e.g., 剂 *ji* 'dose', 宗 *zong*, 桩 *zhuang*, 套 *tao* 'set'), measure classifiers (e.g., 米 *mi* 'meter', 亩 *mu*, 斤 *jin*), container quantifiers (e.g., 杯 *bei* 'cup', 盆 *pen* 'pot', 包 *bao* 'bag'), shape classifiers (e.g., 团 *tuan* 'ball', 块 *kuai* 'piece', 片 *pian* 'slice'), etc. Among auxiliaries, there are structural particles (e.g., 的 *de*, 地 *de*, 得 *de*), aspect markers (e.g., 着 *zhe*, 了 *le*, 过 *guo*), numeral particles (e.g., 来 *lai* 'approximately', 余 *yu* 'over'), etc. The details are omitted to save space here.

5.4 Overlapping of Classes

5.4.1 Lack of Correspondence between Meaning and Distribution in Words

In the logical sense, parts of speech can be seen as the subsets of a whole set of words. These subsets are to be internally consistent or homogeneous, mutually exclusive, and collectively exhaustive. Unfortunately, this is only possible in the ideal situation, as the connection between meaning and distribution in words is very complicated.

Ideally, words exhibit consistent behaviors in both meaning and distribution. For instance, the word 葡萄 *putao* 'grape' denotes an entity in the external world and therefore connects with the idea of such an entity in a human mind. On the other hand, the word combines with other words to form a larger linguistic unit, and hence to express more complex ideas, such as 吃_葡萄 *chi_putao* 'to eat grapes' and 葡萄_涨_价_了 *putao_zhang_jia_le* 'grapes are sold at a higher price', where the word constitutes the object of 吃 *chi* 'to eat' and the subject of 涨_价 *zhang_jia* 'to be sold at a higher price' respectively. As words in the object and subject positions tend to denote entities, the word 葡萄 *putao* behaves consistently in meaning and distribution to be classified as a noun. Similar consistency between meaning and distribution is found with a large number of words in Chinese, which makes the definitional criteria given in Table 5.2 applicable and viable on the whole. In the use of language, however, new combinations often occur, challenging the conventional pairing between meaning and distribution in words and making it different to define internally consistent and mutually exclusive classes of words.

(4)   a. 这本书的**作者**是很有想法的。
          zhe__ben__shu__de__zuozhe__shi__hen__you__xiangfa__de
          The author of the book is quite innovative.
      b. 这本书的**出版**是很有意义的。
          zhe__ben__shu__de__chuban__shi__hen__you__yiyi__de
          The publication of this book is of much significance.

Semantically, 作者 *zuozhe* 'author' and 出版 *chuban* 'to publish' belong to different categories, the former used to denote a person and the latter used to describe an action. In (4), however, the two words occur in the same position of the Attributive-Head construction, which usually requires a noun rather than a verb. Such being the case, 出版 *chuban* can be identified as a noun based on its syntactic property and as a verb based on its semantic property, leading to an overlapping of two classes, namely noun and verb, and the word therefore has multiple memberships.[11] Here are more examples.

(5)   a. 你写你的，我写我的，各抒己见。
          ni__xie__nide__wo__xie__wode__geshujijian
          You write yours and I write mine. Let's each air our own opinion.
      b. 我胖我的，关你什么事？
          wo__pang__wo__de__guan__ni__shenme__shi
          Yes, I'm fat, but is it your business?

---

[11] Such an overlap is avoided in English, as an action can be referred to by a noun or a gerund when functioning as the subject or object of a sentence. For instance, *the publication of the book is too late*, where the action 'to prepare and issue ... for public sale' is expressed by the noun *publication*, while the verbal form of the same action is *to publish*, as in the sentence *the book is published*.

(6)   a.   这种赛制明显是极度不合理的，太容易被田忌了。

       zhe_zhong_saizhi_mingxian_shi_jidu_bu_heli_de_tai_rongyi_bei_tianji_le

       The competition system is extremely unreasonable, with conspicuous loopholes to cash in on.

       b.   汉字整容，我也 "被67%" 了。

       hanzi__zhengrong__wo__ye__bei__67%_le

       As to the readjusted Chinese characters, I was somehow counted in the "67% of the public" claimed to stand for them by the survey.

(7)   a.   他一个人在那里跑来跑去。

       ta__yi__ge__ren__zai__nali__pao__lai__pao__qu

       He is running back and forth on his own over there.

       b.   你这么客气来客气去的，可不算是朋友了。

       ni__zheme__keqi__lai__keqi__qu__de__ke__bu__suanshi__pengyou__le

       You are being too polite. That's not what a friend does.

       c.   你不要啊来啊去的，没完没了。

       ni__buyao__a__lai__a__qu__de__meiwan__meiliao

       Cut it out. Stop your endless ah-this-ah-that.

(8)   a.   那位日本游客 "哈依哈依" 地直点头。

       na__wei__riben__youke__hayi__hayi__de__zhi__diantou

       The Japanese tourist kept nodding while repeating hai-hai.

       b.   你傻呀，别老大姐大姐的，要叫就叫姐。

       na__sha__ya__bie__lao__dajie__dajie__de__yao__jiao__jiu__jiao__jie

       You fool! Stop calling me Elder Sister. Call me Sis, if you like.

(9)   a.   自觉用 "四个全面" 战略布局统一思想行动。

       zijue__yong__si__ge__quanmian__zhanlüe__buju__tongyi__sixiang__xingdong

       Reconcile our views and actions by adopting on our own initiative the Four-Pronged Comprehensive Strategy?

       b.   交通事故后「三不一没有」的处理方式靠谱吗？

       jiaotong__shigu__hou__san__bu__yi__meiyou__de__chuli__fangshi__kaopu__ma

       In case of a car accident, does it work well to adopt the "Three Don'ts and One No" policy?

In the construction Pronoun-X-Pronoun-的, X is often a verb, which is the case with (5a). In (5b), however, the word 胖 *pang* 'fat' occurs in that position, with its adjectival meaning as usual.

In (6a), the proper noun 田忌 *tianji* is placed after 被 *bei*, a position often taken by transitive verbs like 批评 *piping* 'to criticize', 推倒 *tuidao* 'to push down', and 打伤 *dashang* 'to beat and injure'. In (6b), the same position is taken by the numeral '67%'.[12]

---

[12] In (6a), 被_田忌 *bei_tianji* means 'to be defeated with the strategy that Tian Ji once adopted in the horse race', namely, the strategy of using one's own strengths to gain the greatest advantage, where the noun 田忌 *tianji* is just like a transitive verb in its conventional passive usage in a 被-X construction. By contrast, (6b) presents a more recent

In the construction X-来-X-去, which describes the repetition of a movement or an action, X is usually a verb, as is the case with (7a). But in (7b), X is the adjective 客气 *keqi* 'polite', and in (7c), it is the sentence-final particle or interjection 啊 *a* 'ah'.

The expression 哈依 *hayi* 'yes' in (8a) is a transliteration of a Japanese word used to give an affirmative response, and 大姐 *dajie* 'big sister' in (8b) is a noun used to address a woman about one's own age. Both take the position of X in the reduplicated form X-X-的 (Liu 刘丹青 2009)

Modified by 四_个 *si_ge* 'four', the adjective 全面 *quanmian* 'comprehensive' in (9a) plays the Head in the Numeral-Classifier-Head construction, a position usually taken by a noun. In (9a), 不 *bu* 'not' and 没有 *meiyou* 'no' are modified by numerals, which is definitely not the common usage of adverbs. From a discourse point of view, the adjective and adverbs in these examples, when modified by numerals or numeral-classifier phrases, are abbreviations and function as pronouns to refer to the related expressions in their complete forms.[13]

As can be seen from all the examples above, a certain position in a construction originally allows words of class A only, but words of class B are actually found there, bearing their conventional meaning. With the latter, the changed distribution, and hence the lack of correspondence between their meaning and distribution, poses a question: Do these words belong exclusively to class B, or do they belong to both classes?

5.4.2 Handling the Overlapping of Classes

As a matter of fact, the ideal situation described in Section 5.4.1 for word classification requires a huge increase of parts of speech, as the strict pairing of meaning and distribution in words can only be achieved by adding new classes. For instance, the position of 田忌 *tianji* in (6a) sets the word apart from nouns and verbs in general. Thus a new part of speech is required for such words taking the position X in the construction 被-X, and among them is the expression '67%' in (6b). Consequently, thousands of parts of speech have to be defined. (Chen 陈小荷 1999; Zhan 詹卫东 2009; Zhan 詹卫东 2013). Otherwise, there will be a huge number of words that have multiple membership, with 田忌 *tianji* being both a noun and a verb, and '67%' being both a numeral and a verb, which is definitely not a desirable situation in the classification of words.

Rather, a reasonable number of classes are expected, making it possible and convenient to refer to each class, on the one hand, and to distinguish between classes at the appropriate level of granularity, on the other. Thus

---

usage of the 被-X construction, where X can be an intransitive verb, an adjective, a noun, or even a numeral. This construction is used to express that something is claimed to be true while it is not or at least not certain.

[13] The Four-Pronged Comprehensive Strategy stands for finishing building a moderately prosperous society, deepening reform, advancing the law-based governance of China, and strengthening Party self-discipline. The "Three Don'ts and One No" policy stands for "don't make advance payment, don't pay visit to the injured, don't seek mediation personally, and claim to have no cash"?

we need to be well aware of two essential facts when handling words that may belong to more than one class.

First, there are prototypical and non-prototypical usages of the words in a class, just as there are prototypical and non-prototypical members in this class. For instance, the word 出版 *chuban* 'to publish' describes an action, from which its prototypical usages are derived, namely acting as a predicate, taking an object, or being followed by an aspect marker, etc. In (4b), 这本书的出版 is a non-prototypical usage of the word. It is not common, and is mainly found in formal written texts rather than spoken Chinese. It has a constraint that is not observed when the word is used prototypically, namely the word 出版 *chuban* in this usage can be modified by only a few adverbs, such as 这本书的不出版 *zhe_ben_shu_de_bu_chuban* 'the fact that the book is not published', rather than *这些书的都出版 *zhe_ben_shu_de_dou_chuban* 'the fact that these books are all published'. Likewise, 田忌 *tianji* is prototypically and therefore more commonly used as a noun, and '67%' as a numeral. These words are to be classified according to their prototypical usages. The complexity and difficulty of word classification will be unnecessarily increased if non-prototypical usages of words are considered as the definitional criteria.

Second, there are conventional phrasal constructions and unconventional ones, in which the positions of a word and the corresponding constraints can be used to define its part of speech. A conventional construction, as those listed in Table 5.1, allows its constituent positions to be taken by either words or phrases. In other words, the conventional construction is expandable or even recursive. For instance, in 老师_非常_喜欢_这_两_所_大学 *laoshi_feichang_xihuan_zhe_liang_suo_daxue* 'the teacher likes the two universities very much', a Subject-Predicate construction is expanded to include other three phrasal types; and in 学习_写_论文 *xuexi_xie_lunwen* 'to learn to write a paper', there is a recursion of the Verb-Object construction. By contrast, unconventional constructions, such as Pronoun-X-Pronoun-的 in (5), 被-X in (6), and X-来-X-去 in (7), usually require X to be a word, with no expansion, not to mention recursion. These constructions are mainly used with certain rhetorical implications and are therefore not common. Such being the case, the positions of words in unconventional phrasal constructions cannot be taken as the definitional criteria of their parts of speech.

However, it is worth noting that the non-prototypical usages of a word and its distributions in unconventional phrasal constructions do reflect the properties of the word in a more fine-grained way and therefore require appropriate attention. For instance, when the word 出版 *chuban* 'to publish' is identified as a verb, it is necessary to specify that the word can also be the Head in an Attributive-Head construction, a function usually performed by a noun. Likewise, when the word 胖 *pang* 'fat' is identified as an adjective, it is necessary to specify that the word can also take the position of X in Pronoun-X-Pronoun-的.

Feature structures can be used to specify such fine-grained properties of individual words, giving a detailed and flexible account of linguistic facts. As a result, unnecessary changes can be avoided to the definitions of parts of speech and the classification scheme as a whole. Yu et al. 俞士汶等 (2003) introduced a comprehensive description of the grammatical knowledge of Chinese words in which individual words are classified and their fine-grained grammatical attributes specified.

5.5 Conclusion

The purpose of classifying words into parts of speech is to better represent how they combine with each other. For a human, a word is in the first place a means of expressing a concept, based on which some of its usages will be learned. However, it is often impossible for him or her to know all the usages of words, and his or her understanding of word usages may also vary with time. Such being the case, words cannot actually be classified in a strict sense, with all their distributions considered. Instead, it is more reasonable and feasible to group them into coarse-grained classes according to their meanings and their prototypical usages in conventional constructions, and then describe their non-prototypical usages, as individual words, in unconventional constructions, using feature structures or other formalisms.

The discussion about the parts of speech in Chinese in this chapter takes a structural point of view, from which definitional criteria are given for each class, together with a solution to deal with the overlapping of classes. It is hoped that the discussion may shed light on lexicography, Chinese language teaching, and Chinese language processing.

Studies on the classification of words in Chinese also take the perspectives of generative linguistics, functional linguistics, linguistic typology, etc. (Huang et al. 2009; Li and Thompson 1981; Li et al.1984; Huang et al. 黄昌宁等 2009; Yuan et al. 袁毓林等 2009; Yuan 袁毓林 2010; Guo 郭锐 2012; Shen 沈家煊 2007, 2015; Xiong 熊仲儒 2016) In the field of Chinese language processing, part-of-speech tags are defined to label words in electronic dictionaries and corpora. Careful observations and inspiring analysis have been made accordingly. For more details, see Huang et al. 黄昌宁等 (2009), Huang and Li 黄昌宁, 李玉梅 (2009), Song 宋柔 (2009), Song and Xing 宋柔, 邢富坤 (2009), etc.

References

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese.* Berkeley, CA: University of California Press.

Chen, Xiaohe 陈小荷. 1999. Parts-of-speech in Chinese – From the perspective of automatic syntactic parsing 从自动句法分析角度看汉语词类问题. *Language Teaching and Linguistic Studies* 语言教学与研究 3:63–72.

Guo, Rui 郭锐. 2002. *A study on the parts-of-speech in contemporary Chinese* 现代汉语词类研究. Beijing: Commercial Press.

Guo, Rui 郭锐. 2012. Typology of adjective and the grammatical status of adjective in Mandarin Chinese 形容词的类型学和汉语形容词的语法地位. *Chinese Language Learning* 汉语学习 5:3–16.

Guo, Rui 郭锐. 2014. Chapter 4.2 in *Contemporary Chinese* 现代汉语第 4.2节, ed. Yang Shen and Rui Guo 沈阳, 郭锐. Beijing: Higher Education Press.

Huang, C.-T. James, Y.-H. Audrey Li, and Yafei Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University Press.

Huang, Changning, Zixia Jiang, and Yumei Li 黄昌宁, 姜自霞, 李玉梅. 2009. Verb modified by adjective directly: An ambiguous construction of "a+v" 形容词直接修饰动词的 "a + v" 结构歧义. *Studies of the Chinese Language* 中国语文 1: 54–63.

Huang, Changning, and Yumei Li 黄昌宁, 李玉梅. 2009. Evaluation of the four solutions to the problem of verb-noun overlapping words: Observations from language engineering 评动、名兼类词的四种划分策略——来自语言工程的观察. *Journal of Chinese Linguistics* 语言学论丛 40:74–92.

Huang, Chu-Ren and Dingxu Shi. (eds.) 2016. *A reference grammar of Chinese*. Cambridge: Cambridge University Press.

Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

Li, Y. C., Robert L. Cheng, Larry Foster, Shang H. Ho, John Y. Hou, and Moira Yip. 1984. *Mandarin Chinese: A practical reference grammar for students and teachers*. Taipei: Crane Publishing.

Liu, Danqing 刘丹青. 2009. Ideophonic reduplication of content words and the related constructions in Mandarin Chinese 实词的拟声化重叠及其相关构式. *Studies of the Chinese Language* 中国语文 1:22–31.

Lu, Jianming 陆俭明. 1994. On the overlapping of word classes 关于词的兼类问题. *Studies of the Chinese Language* 中国语文 1:28–34.

Lu, Jianming 陆俭明. 2005. *A course on the grammatical study of contemporary Chinese* 现代汉语语法研究教程, third ed. Beijing: Peking University Press.

Shen, Jiaxuan 沈家煊. 2007. Nouns and verbs in Chinese 汉语里的名词和动词. *Sino-Tibetan languages* 汉藏语学报 1:27–47.

Shen, Jiaxuan 沈家煊. 2015. Word class typology and the Chinese nominalism 词类的类型学与汉语的词类. *Contemporary Linguistics* 当代语言学 2:127–145.

Shi, Dingxu 石定栩. 2009. Problems in the classification of words in Chinese 汉语词类划分的若干问题. *Journal of Chinese Linguistics* 语言学论丛 40:93–110.

Song, Rou 宋柔. 2009. Parts-of-speech in Chinese from the perspective of language engineering 从语言工程看汉语词类. *Journal of Chinese Linguistics* 语言学论丛 40:23–38.

Song, Rou, and Fukun Xing 宋柔, 邢富坤. 2009. Essentials and difficulties in the classification of words in Chinese from the perspective of language engineering 从语言工程看现行汉语词类体系的本质和困境. In *Symposiums on new visions for the parts-of-speech in Chinese* "新视野下的汉语词类问题"系列研讨.

Xiong, Zhongru 熊仲儒. 2016. Study of Chinese categories from perspective of Generative Grammar 从生成语法看汉语词类研究. *Journal of Anhui Normal University (Hum. & Soc. Sci.)* 安徽师范大学学报（人文社会科学版）4:428–432.

Xu, Shu and Jingchun Tan 徐枢, 谭景春. 2006. Remarks on the part-of-speech tagging in the Contemporary Chinese Dictionary (5th ed.) 关于《现代汉语词典（第 5 版）》词类标注的说明. *Studies of the Chinese Language* 中国语文 1:74–86.

Yu, Shiwen, Xuefeng Zhu, Hui Wang, Huarui Zhang, Yunyun Zhang, Dexi Zhu, Jianming Lu, and Rui Guo 俞士汶，朱学锋, 王惠, 张化瑞, 张芸芸, 朱德熙, 陆俭明, 郭锐. 2003. *The grammatical knowledge-base of contemporary Chinese: A complete specification* 现代汉语语法信息词典详解, second ed. Beijing: Tsinghua University Press.

Yuan, Yulin, Hui Ma, Ren Zhou, and Hong Cao 袁毓林, 马辉, 周韧, 曹宏. 2009. *A handbook of word classification in Chinese* 汉语词类划分手册. Beijing: Beijing Language and Culture University Press.

Yuan, Yulin 袁毓林. 2010. *A cognitive investigation and fuzzy classification of word-class in Mandarin Chinese* 汉语词类的认知研究和模糊划分. Shanghai: Shanghai Educational Publishing House.

Zhan, Weidong 詹卫东. 2000. *A study of construction rules in contemporary Chinese for Chinese information processing* 面向中文信息处理的现代汉语短语结构规则研究. Beijing: Tsinghua University Press.

Zhan, Weidong 詹卫东. 2009. A syntactic perspective on the classification of words in Chinese: Three basic questions 词类三问——一个汉语词类知识学习者和使用者的反思. *Journal of Chinese Linguistics* 语言学论丛 40:66–73.

Zhan, Weidong 詹卫东. 2013. A review on word classification of Chinese from the perspective of sentence parsing by computer 计算机句法结构分析需要什么样的词类知识——兼评近年来汉语词类研究的新进展. *Studies of the Chinese Language* 中国语文 2:178–190.

Zhu, Dexi 朱德熙. 1982. *Lecture notes on grammar* 语法讲义. Beijing: Commercial Press.