

# A root-and-pattern approach to word-formation in Chinese

Cheng Gong & Liu Ying  
Zhejiang University

This paper argues that compounding, the major source of word-formation in Chinese, and the root-and-pattern system in Hebrew involve fundamentally the same syntactic operations and observe the same locality constraints, despite the salient differences. More specifically, it addresses the well-known continuum that the coordinate and attributive compounds behave more like words, whereas resultative and subordinate compounds are much more like phrases. It puts forward the idea that this continuum can be accounted for by assuming that there is a distinction between word-formation from roots and word-formation from words, with the former giving rise to more lexical properties and the latter more phrasal properties. The paper also discusses some related issues, such as the correct formulation of word-level phases and the structure of the major types of compound words in Chinese.

**Keywords:** root-and-pattern, root, Chinese, compounding, Distributed Morphology

## 1. Introduction

Of the languages spoken in Asia, Chinese and Hebrew seem to be separated by a chasm that contains glaring differences in almost every respect, with word-formation probably being the most striking. While Chinese resorts mostly to compounding, as will be discussed below, Hebrew has a unique way of building words known as the ‘root-and-pattern’ system. Roots, often represented by the symbol  $\sqrt{\quad}$ , are the lexical kernel in the form of a set of segmental consonants arranged in a specific sequence, with the following properties: (i) they are capable of producing words of any category (nouns, verbs, or adjectives), depending on the specific syntactic environment into which they are placed; (ii) they are unpronounceable by themselves as well as underspecified semantically, open to multiple interpretations. By contrast, patterns, consisting of a set of interlocking vowels, function

to add grammatical information to the roots, in particular the categorial feature. They also make the consonantal sequence pronounceable and fix its semantic interpretation, sometimes in combination with prefixes or suffixes. In other words, a Hebrew root is not an actual word and becomes one only when merged with a pattern. As a concrete example, the Hebrew root  $\sqrt{\text{sgr}}$  may be placed in different patterns and turned into nouns and verbs that are somewhat loosely related in meaning. This is illustrated below (adapted from Arad 2003: 746. Her (10)):

- (1)  $\sqrt{\text{sgr}}$
- a. CaCaC (v) *sagar* v, 'close'
  - b. hiCCiC (v) *hisgir* v, 'extradite'
  - c. hitCaCCeC (v) *histager* v, 'cocoon oneself'
  - d. CeCeC (n) *seger* n, 'closure'
  - e. CoCCayim (n) *sograyim* n, 'parentheses'
  - f. miCCeCet (n) *misgeret* n, 'frame'

When added to the vowel pattern  $-a-a-\emptyset$ , as in (1a), the triconsonantal root  $\sqrt{\text{sgr}}$  yields a verb *sagar*, which means 'close', and when added to the pattern  $e-e-\emptyset$ , produces the noun *seger*, meaning 'closure', as in (1d).

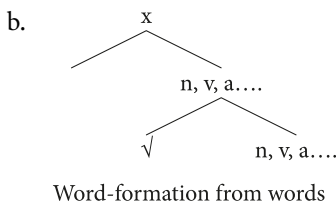
In a series of works, Arad (2003, 2005) develops a syntactic account for the root-and-pattern method of word-formation in Hebrew in the framework of Distributed Morphology (DM). Three important premises underlie her account. First, roots are atomic lexical elements devoid of any syntactic or functional material. This, in turn, means that roots are category-neutral and are capable of becoming a noun or a verb depending on whether they are inserted in a nominal or verbal environment. Patterns, on the other hand, contain the functional material needed to turn the roots into words. Moreover, the combination of the root with the pattern is regulated by locality constraints such that each Merge yields a phase in the sense of Chomsky (1998), that is, a stage in the derivation where the element built by the computational system is spelled out both semantically and phonologically. A particularly interesting consequence of such a conception of word-formation is that it predicts a crucial distinction between word-formation from roots and word-formation from existing words shown below.

- (2) a.
- 
- ```

      x
     / \
  √      x

```

Word-formation from roots



In (2a), a root is merged with a functional head ( $x$ ) bearing a categorial feature,  $n$ ,  $v$ , or  $a$ , and is turned into a word. As a phase head,  $x$  defines a closed domain for assigning multiple interpretations to the root, depending on the environment; which cannot be predicted from the combination of the root and the categorizing head. In (2b), by contrast, the head  $x$  is merged above this word and, due to the working of the phase, cannot ‘see’ the root below it whose interpretation has already been fixed. The interpretation of the phrase headed by  $x$  in (2b) is tightly restricted by properties of the source words, exactly as predicted by the phase theory. When (3) is considered as an example, Arad shows that a new Hebrew verb, *misger* ‘to frame’ may be formed from the noun *misgeret* (‘frame’, see (1f)) when placed in the verbal pattern CiCCeC, as in (3b):

- (3)  $\sqrt{sgr}$   
 a. miCCeCet *misgeret* ‘a frame’  
 b. CiCCeC *misger* ‘to frame’

Arad also argues that this noun-derived verb differs from root-derived verbs not only morphologically, but also semantically. That is, while the root  $\sqrt{sgr}$  may be assigned numerous interpretations, *misger* seems to be tied to the meaning of *misgeret*, from which it derives.

A great strength in Arad’s account lies in its more nuanced and fine-grained treatment of the root-and-pattern method in Hebrew. Instead of treating it as a monolithic system, Arad successfully demonstrates that different local domains are operative in the system, such that idiosyncracies may be exhibited in the domain where a root is merged to a functional head; semantically more transparent formations may be derived in the domain where a functional affix merges with an existing word whose interpretation is already fixed in a given environment.

The present paper is intended to treat compounding in Chinese with essentially the same fine-grained approach that Arad adopts for the Hebrew root-and-pattern system. More specifically, it focuses on the well-known continuum that extends from more lexical to more phrasal properties among Chinese compounds. Here coordinate and attributive compounds behave more like words, whereas resultative and subordinate compounds behave much more like phrases. We will argue that viewing compounding in Chinese as being on a par with the

root-and-pattern system in Hebrew could help explain the hitherto mysterious continuum. Indeed, like the root-and-pattern system in Hebrew, Chinese compounding also involves a process that manipulates different positions in the syntactic tree in tandem with locality constraints. In this way, the manifested lexical properties are derived in the inner domain with roots as input, whereas phrasal properties come from the outer domain with words as the input.

Some caveats are requested before a more detailed discussion is undertaken. The scope of our data is confined to disyllabic primary compounds, the most common and typical form of what is identified as words in Chinese.<sup>1</sup> According to Zhou (2005:110), 32,346 dissyllabic compounds in total are collected in the *Xiandai Hanyu Cidian* 'Dictionary of Modern Chinese' (DMC), the best known and highly reputable dictionary of Chinese, accounting for 57.8% of all its entries. In fact, we will focus on the typical cases in the formation of disyllabic compounds and consider those involving more complicated derivational processes as beyond the scope of this paper. Moreover, for the purposes of the present study, we shall not cover all types of compounds, but instead limit the discussion to four major groups, the coordinate, attributive, resultative, and subordinate. Definitions of these groups are provided in the subsections to follow. According to Zhou's (2005) study, these four types account for 94.65% of all the compounds listed in the DMC, and therefore enough to be considered as comprehensive in terms of empirical coverage.<sup>2</sup>

---

1. Other devices such as affixation and reduplication are also used but account for a much less significant proportion when compared with compounding.

2. Another common type, the so-called subject-predicate compounds, are illustrated below:

- a. di-zhen  
earth-quake  
'earthquake'
- b. tou-teng  
head-ache  
'headache'
- c. guo-ying  
state-own  
'state-owned'

Compounds similar to those in (i) also abound in languages such as English. This type is left untreated in the present paper primarily for two reasons. First, they may be more complex than appearances might suggest. Indeed, what has traditionally been identified as *predicate* in these compounds may be pure nominals, whereas others may be participial constructions, a fact that is suggested by the English translation. A detailed analysis would lead the discussion too far afield. Second, they are relatively less productive, contributing only 1% of all compounds. For these reasons, we exclude the subject-predicate compounds from the present discussion to avoid unnecessary complications (see Yang Zhou 2015 for a detailed review).

The rest of the paper is organized as follows. Section 2 is devoted to establishing the continuum from the more lexical to more phrasal properties among the major types of Chinese compounds. Section 3 provides a discussion of the proper formulation of xP phases. Section 4 explicates a new model of the underlying structures and derivations associated with the types of compounds. Section 5 introduces further evidence for the proposed revision to the notion of phase. Section 6 concludes the article.

## 2. The continuum of lexicality in Chinese compound words

This section is devoted to establishing the existence of a continuum of lexicality among Chinese compound words. The term *lexicality* is used here to mean properties that are typically attributed to words (rather than phrases). We use the following set of criteria in order to identify the structural properties of the different types of compound, adapted from various sources in existing literature:<sup>3</sup>

- (4) a.  $\pm$ endocentric
- b.  $\pm$ compositional
- c.  $\pm$ interruptible

In (4), the first criterion, [ $\pm$ endocentric], predicts lexicality from the categorial status. It suggests that a structure behaves more like a lexical word if its category is different from any of the constituents, a scenario known as ‘exocentric’ or [-endocentric]; otherwise, it is more phrasal. The second criterion, [ $\pm$ compositional], relates lexicality to semantic behavior; it suggests that a structure whose meaning cannot be predicted from either of the constituents is more lexical; otherwise it is more phrasal. The third criterion, [ $\pm$ interruptible], is based on morphological properties; a structure is considered to be more lexical if it does not allow the separation of its parts by an extraneous element; otherwise it is more phrasal.

---

For an account of synthetic compounds in the Distributed Morphology framework, please refer to Cheng and Zhou (2015).

3. The criteria in (4) are adapted from proposals for whether a sequence should be identified as a word or a phrase in existing literature. For instance, Chao (1968) proposed a comprehensive set of criteria which include the following: (a) whether part of the item is a bound form; (b) whether part of it is neutral-toned; (c) whether the meaning of the whole is compositional of its parts; (d) whether the internal structure is exocentric; (e) whether the parts are inseparable from each other. If the answer to these criteria is positive, it is a word; otherwise it is a phrase (see also Huang 1984 for a related discussion). As can be seen, the present account adopts his (c), (d), and (e), which we believe to be sufficient for the present purposes.

With the aid of the criteria in (4), we will examine the structural properties of the four major types of Chinese compounds under different headings in the rest of this section, beginning with coordinates.

## 2.1 Coordinates

Coordinate compounds are formations whose constituents are structurally equal, with neither dominating the other (see, for instance, Bisetto and Scalise 2005 for a discussion). In Chinese, they are typically a union of two (or more) elements with either a synonymous meaning, as in (5a), or an antonymous meaning, as in (5b):

- |     |    |                 |               |                 |
|-----|----|-----------------|---------------|-----------------|
| (5) | a. | yu-yan          | dou-zheng     | feng-fu         |
|     |    | language-speech | fight-compete | plenty-rich     |
|     |    | 'language'      | 'fight'       | 'plenty'        |
|     | b. | kai-guan        | hu-xi         | mao-dun         |
|     |    | open-close      | inhale-exhale | spear-shield    |
|     |    | 'switch'        | 'breathe'     | 'contradictory' |

Coordinates are generally believed to have the following properties when tested against the criteria in (4). First, their syntactic categories cannot be reliably predicted from their component constituents. That is, though many coordinate compounds are endocentric in the sense that their category is the same as that of a constituent, as shown in (5a), a fairly large proportion have a category that cannot be inferred from one of their constituents, especially for those antonymous pairings, as shown in (5b).

A particularly clear example is given below. As we can see, each item in the group contains *ji* 'surge' and another verb-like morpheme, yet (6a) and (6b) turn out to be adjectives whereas (6c) is a verb.

- |     |    |               |
|-----|----|---------------|
| (6) | a. | ji-dong (a.)  |
|     |    | surge-move    |
|     |    | 'excited'     |
|     | b. | ji-lie (a.)   |
|     |    | surge-intense |
|     |    | 'fierce'      |
|     | c. | ji-li (v.)    |
|     |    | surge-sharpen |
|     |    | 'encourage'   |

Further support for the argument that coordinates are low in endocentricity comes from recognition that they are the major source for disyllabic adverbs, a relatively minor category. Examples like (7), also discussed in Zhang (2007), illus-

trate that adverbs are frequently derived from the merging of morphemes that are typically considered as nominal, adjectival, or numeral.

- (7) a. fan-zheng  
back-front  
'anyway'
- b. hao-dai  
good-bad  
'anyway'
- c. qian-wan  
thousand-ten thousand  
'surely'

The second well-known property of coordinate compounds is that they show a high degree of semantic non-compositionality, a lexical rather than phrasal property according to (4b). Many of the instances listed above, especially those in (6)–(7) illustrate that the meaning of a coordinate cannot be reliably predicted on the basis of its constituents. Instances in (8)–(9) below suggest a different but closely related property, namely that the roots in coordinates have a common core of meaning yet to be fixed in actual words, a property that closely resembles those in Hebrew.

- (8) a. feng-yu  
wind-rain  
'wind-and-rain'
- b. feng-shui  
wind-water  
'geomancy'
- c. feng-yue  
wind-moon  
'romance'
- (9) a. gu-ge  
human bone-animal bone  
'skeleton'
- b. gu-gan  
bone-stem  
'backbone'
- c. gu-rou  
bone-flesh  
'kindred'

In (8) the root *feng* literally means 'wind', yet words built around it differ widely in meaning. As suggested in the translation, apart from metaphorical extensions,

*feng-yu* ‘wind and rain’ in (8a) denotes a concrete meaning, while *feng-shui* ‘geomancy’ and *feng-yue* ‘romance’ in (8b) and (8c) are more abstract, and their relation to the wind is indirect and remote. More or less the same is true of (9), where only (9a) is directly linked to biological bones, while (9b) is most likely to refer to someone who is the backbone of a group, and (9c) a relation by blood.

Another feature that may be interpreted as a sign of semantic non-compositionality is that coordinates sometimes take the meaning of just one of the morphemes, a phenomenon known as ‘meaning loss’ (see Cao 2016 for review). In the following instances, the compounds assume the meaning of the first morpheme and are entirely oblivious to that of the second:

- (10) a. guo-jia  
country-home  
‘country’  
b. chuang-hu  
window-door  
‘window’  
c. ren-wu  
man-thing  
‘man’

A test against the third criterion confirms that coordinates are associated with strong lexical properties in that they are non-interruptible. If, for instance, we insert a conjunction *he* ‘and’, *yu* ‘and’, or *huo* ‘or’, between the two morphemes, the result is either unacceptable in some, such as \**mei-he-li* ‘beautiful-and-fair’, \**ji-yu-lie* ‘surge-and-intense’, or become phrases with an accompanying change in meaning in others, like *hao he dai* (meaning ‘good and bad’), *gu he rou* (meaning ‘bone and flesh’). In any case, the conclusion seems to be that there is almost no way for coordinates to be separated by any extraneous element.

To sum up: the test against the criteria stated in (4) suggests that coordinate compounds are frequently exocentric rather than endocentric in syntactic categories, that they tend to be non-compositional in meaning, and are uninterruptible. In view of such evidence, it seems plausible to rank them as high on the scale of lexical properties.

## 2.2 Attributive compounds

As described in Bisetto and Scalise (2005: 327), attributive compounds are ‘formed either by a noun and an adjective, as in *blue cheese* (where the adjective expresses a property and is in a modifier relation to the noun) or by two nouns, where the



non-head very often is used metaphorically, expressing an attribute of the head (cf. *snail mail*, *sword fish*).<sup>4</sup> The following are some examples from Chinese:<sup>4</sup>

- (11) a. xue-sheng    xue-zhe    xue-shi  
          learn-person   learn-person   learn-man  
          ‘student’    ‘scholar’    ‘bachelor’s degree’  
       b. da-ren    xiao-ren    hong-ren  
          big-person   little-person   red-person  
          ‘adult’    ‘villain’    ‘popular person’

Attributives are generally considered to possess the following properties. First, unlike coordinates, they are typically endocentric; in other words, the grammatical category of the compound as a whole is the same as one of its constituents. In (11), all instances are nominal, the same as the second morpheme, sometimes (mistakenly) dubbed the *head*; the first morpheme is correspondingly called the *non-head*. Second, like coordinates, attributive compounds are highly non-compositional and prone to idiomaticity. For instance, there is no logical reason why a *xue-zhe* ‘learn person’ should mean an academically accomplished person, whereas *xue-shi* ‘learn man’ refers to a bachelor’s degree, given that *zhe* and *shi* share a very similar meaning (see the gloss in (11a)). Similarly, we cannot infer from the literal meaning of the morpheme that *da-ren* ‘big person’ means an adult, as opposed to *xiao-ren* ‘little person’, which means a villain. Finally, like coordinates, they are internally very cohesive, disallowing any syntactic operation on their subparts, like deletion and coordination, let alone movement.

In addition to the aforementioned general properties, attributives exhibit some notable peculiarities on the part of the modifier. Foremost among these is the fact that verb-like elements are capable of modifying the noun-like head without any morphological change. Thus, *xue* ‘learn’, typically regarded as a verb in other contexts, directly collocates with a noun in (11). Such instances abound in Chinese, as illustrated below:

- (12) pao-xie            fei-ji                    kao-rou  
       run-shoe       fly-machine       roast-meat  
       ‘running shoe’ ‘flying machine (plane)’ ‘roasted meat’

---

4. For the purpose of this paper, we confine our discussion to modifier-noun structures and skip the parallel modifier-verb structures, which consist of an adverb-like morpheme as their first element and a verb-like morpheme as the second, as shown below:

- (i) chi-dao    zhong-shi    an-sha    shen-ru  
      late-come   serious-look   secret-kill   deep-enter  
      ‘arrive late’ ‘take seriously’ ‘assassinate’ ‘penetrate’

By contrast, a verb-like element in English is typically marked with a participle when it modifies a noun head, as indicated in the translation in (12). It is intriguing that attributives in the two languages should differ in this respect. We will return to this point in Section 4.3.

Another well-known fact is that some modifiers seem to have no apparent semantic contribution whatsoever. The addition of such an informationally redundant syllable is generally considered as a strategy to meet phonological/prosodic requirements (see Feng 1997 for a detailed discussion). The following are some examples:

- (13) xian-yan shui-jing shang-dian zhu-jia  
 salty-salt water-well trade-store live-house  
 'salt' 'well' 'store' 'house'

Moreover, unique morphemes like *cran-* in *cranberry* are not a particularly rare occurrence in Chinese attributive compounds. Instances such as (14) are frequently found in the names of fruits and vegetables:

- (14) ping-guo bo-cai qiao-mai wan-dou  
 ping-fruit bo-vegetable qiao-wheat 'buckwheat'  
 'apple' 'spinach' wan-bean 'peas'

The first morpheme in each of the words on the list is not found anywhere else in Chinese, an indication of a unique morpheme (cf. Bloomfield 1933:160). Similar instances also occur in the names of cities, rivers, and mountains (see Yang 2003, Chapter 2, among others).

In this paper, we take the facts illustrated in Examples (12)–(14) as evidence that most modifiers in the attributives are roots.<sup>5</sup> We will return to this point in conjunction with the discussion in Section 4.2.

To sum up: attributive compounds differ from coordinate ones in being more endocentric in grammatical categories, but resemble them with respect to semantic compositionality and non-interruptability. They are therefore less lexical than coordinates. Moreover, modifiers in such a structure frequently manifest properties that are associated with roots, such as the ability of a verb-like element to modify a noun directly, the lack of any semantic contribution, and the use of unique morphemes.

---

5. Non-root modifiers are also possible, as cases of nominals turned from VPs discussed in Section 4.4. We put these cases aside.

### 2.3 Resultative compounds

Resultative compounds, also known as VV compounds, refer to the structure composed of two verb-like morphemes, with the first denoting an activity and the second the resulting event or state (see also Huang 1984). The examples are illustrated below:

- (15) da-po    da-pao    kan-dong    ku-xing  
       hit-break hit-run    look-understand cry-awake  
       ‘break’   ‘make-run’ ‘understand’   ‘wake up (sb.) with a cry’

Resultative compounds raise questions on a large range of issues in Chinese linguistics, both conceptual and empirical (see Li 1990, 1995; Shiao 2015 and references therein). In this paper, we focus on facts related to the criteria in (4). Firstly, these compounds are highly endocentric. They typically behave as verbs, in the same way as their constituent parts, and most have the ability to take a nominal complement. Secondly, the meaning of such compounds is highly transparent, so much so that only a tiny proportion needs to be collected in a dictionary.<sup>6</sup> Finally, they have strong internal cohesion, shown by the fact that aspectual markers such as the perfective *le* and experiential *guo* are barred from appearing between the two morphemes. However, this is not without contention. Some previous studies, including Chao (1968) and Thompson (1973), hold that resultatives allow the limited insertion of an extraneous element between the parts, namely, the insertion of two morphemes denoting the ‘potential’ modality: *-de* ‘be able’, or *-bu* ‘be unable’. *De* and *bu* are therefore identified as infixes, a set that only contains these two members in Chinese. The following sentences are taken from Thompson (1973:361):

- (16) a. ta la- de- kai men  
           he pull- can- open door  
           ‘he can pull the door open.’  
       b. ta la- bu- kai men  
           he pull- can’t- open door  
           ‘he can’t pull the door open.’

Other scholars, however, argue against this analysis of infixation (see Cheng and Yang 2016 for a recent and more detailed discussion). In any case, the general picture seems clear: resultative compounds exhibit strong internal cohesion, allowing either no separation at all or limited separation by one of the only two

6. In MCD, resultatives take up less than 3% of all the disyllabic compounds in Chinese. This proportion is not an accurate reflection of their real productivity.

extraneous elements *-de* and *-bu*, depending on whether the theory allows infixation in Chinese.

To summarize, resultative compounds are typically endocentric, highly compositional in semantics, and practically non-interruptible. Judging by the criteria in (4), they are more phrasal than lexical in nature.

Apart from the aforementioned facts, there are other facts suggesting that resultative compounds have phrasal rather than lexical properties. It is significant that, unlike coordinates and attributives, they are absent in the lexical inventory of most world languages. Rather, similar structures in other languages, say, the familiar European languages, are apparently phrases. For instance, English resultatives are undoubtedly phrases and typically assume a discontinuous order so that the object surfaces between the two verb-like morphemes, as in *he pounded it flat* and *he kicked the plank apart*. In fact, the closest counterparts of resultative compounds in Chinese are widely considered to be the so-called serial verb constructions (SVCs) in many African languages and the Caribbean Creoles (see, for instance, Baker 1989; Collins 1994, 2002; Aikhenvald and Dixon 2006; Williams 2008). As Collins (2002: 3) concludes, both structures can be defined in the same way as ‘a succession of verbs and their complements (if any) with one subject and one tense value that are not separated by any overt marker of coordination or subordination’. Moreover, the same range of meanings is expressed by both constructions (directional, consecutive, benefactive, etc.); in most cases, the same verbs that are used in verbal compounds (e.g., ‘cook’, ‘eat’) can also be used to form SVCs. It, seems plausible, therefore, to analyze resultative compounds as phrasal structures, with a main verb taking a complement verb, in the same way that SVCs are usually analyzed.

More convincing clues seem to come from historical and dialectal data in the Chinese language family itself, as the two verbs in a resultative compound might be separated by a nominal complement. The following instances of verbs with a discontinuous order are from the Shanghai dialect in present-day Chinese. This difference is generally regarded as a result of prosodic requirements, which have no effect whatsoever on semantic interpretation.

(17) Shanghai dialect

- a. sɔ<sup>53</sup> fi<sup>23</sup> su<sup>53</sup>  
     cook it   tender  
     ‘cook it tender.’
- b. kʰɔ<sup>53</sup> fi<sup>23</sup> se<sup>34</sup>  
     crack it   broken  
     ‘crack it to pieces’

There is another fact about resultative compounds that deserves serious attention: the categorial status of V2 is ambiguous in that it is predicative but cannot be reliably diagnosed as a verb or an adjective based on standard tests such as adverbial modification and forms of reduplication, among others. This, we believe, constitutes strong evidence for taking V2 as a root, in line with the framework of the DM.

Taking all the aforementioned facts into consideration, it would seem reasonable to assume that resultative compounds have an underlying phrase-like structure. We will return to this point in Section 4.3.

## 2.4 Subordinate compounds

The last type under discussion is the subordinate compound, also known as the predicate-argument compound, which involves a complement relation between the two constituents, shown as follows:<sup>7</sup>

- (18) a. dan-xin  
         carry-heart  
         ‘worry’  
       b. fu-ze  
         carry-duty  
         ‘be responsible for’  
       c. li-fa  
         arrange-hair  
         ‘have a haircut’

Unlike their counterparts in Indo-European languages, such as *pickpocket* and *cutthroat* in English, or *portacenere* ‘ashtray’ in Italian, subordinate compounding is highly productive in Modern Chinese, accounting for 15.6% of all disyllabic compound words collected in the DMC.<sup>8</sup>

---

7. There are also subordinate compounds without a canonical predicate-complement relation between the two constituents, shown in the following examples:

- (i) yang-shang          wo-chuang    ta-chun          kai-dao  
     recuperate-wound   lie-bed          walk-spring      open-knife  
     ‘heal one’s wounds’ ‘stay in bed’ ‘go on a spring outing’ ‘perform an operation’

Zhou and Cheng (2018) argue that objects in (i) are introduced into the structure by a high applicative morpheme.

8. Unproductive and exocentric subordinate compounds like *pickpocket* also exist in Chinese, such as *bang-tui* (tie-leg) ‘legging’, *ling-shì* (lead-affairs) ‘consul’ and *jie-zhi* (forbid-finger) ‘ring’. Such instances are discussed in Huang (1984) and Packard (2000), among others. These are outside the scope of the current discussion.

As far as structural properties are concerned, there is a near consensus among Chinese grammarians that subordinates are the least lexical of all types of compound (see Dong 2009 and references therein). This assumption is confirmed by tests against the criteria in (4). First, an overwhelming majority have a predictable verbal nature and are therefore endocentric, on a par with subordinate phrases. Some have uses other than verbs, but are nonetheless closely connected with their verbal origins.<sup>9</sup> Second, most subordinates have a high degree of semantic transparency, in that most have an interpretation that parallels those of subordinate phrases. Finally, they have the weakest internal cohesion of all compounds in Chinese, so much so that most allow the insertion of extraneous elements or movement. This phenomenon is known as the ‘separable word’ and is illustrated by the following examples:

- (19) a. nimen dan shenme xin?  
           you carry what heart  
           ‘what are you worrying about?’  
       b. wo bang bu liao nide mang  
           I help not ASP your busy  
           ‘I cannot help you’

Zhou (2005) reports that 65% of subordinate compounds may be used with the two morphemes separated, which should be taken as a sure sign of their phrasal properties. As illustrated by (20) below, cases suggestive of movement are frequently found, a fact that is also discussed by Huang (1984) and many others.

- (20) a. xin, wo yidian dou bu dan  
           heart, I a-bit all not carry.  
           ‘worried, I am not at all.’  
       b. zao, ni xi le ma?  
           bath, you wash ASP Q  
           ‘have you taken a bath?’

Ever since Lu (1957) and Chao (1968) first brought instances like (19) and (20) to our attention, numerous proposals have been put forward on whether subordinates should be treated as words or phrases (see, for instance, Huang 1984; Zhang 2007, among others). We will not go into a detailed discussion, merely noting here that the existence of the so-called separable words offers a glimpse into the shaky

---

9. Jinxue Li (2012) gives the following statistics based on his study of 5491 subordinate compounds in his corpus: 84.05% is verbal, 2.66% nominal, 4.63% adjectival, 0.91% adverbial, and the rest (7.92%) have multiple categorial usages, depending on the context.

distinction between words and phrases on the one hand, and salient phrasal properties among subordinate compounds on the other.

Why should subordinate compounds behave very much like phrases? The widely accepted answer in traditional grammar is that they come from corresponding phrases via a process denoted by the umbrella term *lexicalization* by which an existing material such as a phrase develops or is recruited to form lexical items (Hopper and Traugott 1993: 224 n. 5). It is believed that some originally VO sequences have become so lexicalized that their objects have lost their nominal properties and have gradually been reduced to phonological place-holders without any syntactic role to play. For instance, it is well-known that *chu-ban* 'publish version' and *huai-yi* 'have suspicion', two subordinate compounds in Archaic Chinese, have become disyllabic verbs in Modern Chinese and can each take an argument complement, as shown below:

- (21) a. Zhang jiaoshou chuban le yi-ben shu  
 Zhang professor publish ASP a-CL book  
 'Prof. Zhang had a book published.'  
 b. women huaiyi ta shou le huilu  
 we suspect him take ASP bribe  
 'we suspected him for taking bribes.'

Based on the discussion above, it is clear that subordinate compounds share the most phrasal properties among all the compounds examined so far, being typically endocentric, highly compositional in meaning, and frequently interruptible.

## 2.5 Summary

To sum up the discussion in this section, compound words in Chinese are found to be more heterogeneous than noted in previous literature. Using the four criteria listed in (4) as diagnostics, we examined four types of compound in Chinese. The results are summarized in Table 1, which clearly indicates a continuum from more lexical to more phrasal properties.

**Table 1.** The continuum in Chinese compounds

|                              | Coordinate | Attributive | Resultative | Subordinate |
|------------------------------|------------|-------------|-------------|-------------|
| Endocentric                  | low        | high        | high        | high        |
| Compositional                | low        | low         | high        | high        |
| Interruptible                | low        | low         | low         | High        |
| Continuum: lexical → phrasal |            |             |             |             |

This this continuum suggests that, from the left to the right, Chinese compounds show a decrease in lexical properties and an accompanying increase in phrasal properties. As such, coordinates are the most word-like, followed by attributives. Comparatively speaking, resultatives are more phrasal than the first two but less phrasal than subordinates. In this continuum, subordinate compounds have the most phrase-like properties.<sup>10</sup>

The findings shown in Table 1 are not entirely new; instead, they are implied in numerous studies in existing literature. An interesting question arises if Table 1 accurately characterizes properties of Chinese compounds, namely, what causes such a varied degree of lexicality? More importantly, how should we best capture such a continuum? In the following sections, we shall develop an account of the syntactic approach in line with Arad's account outlined in Section 1.

### 3. Theoretical issues

In the preceding section we endeavored to demonstrate the existence of a continuum of lexicality among the major types of Chinese compounds. Could the continuum, or more generally, could be properties of Chinese word-formation, be accounted for along essentially the same lines, as Arad (2003, 2005) suggests for Hebrew word-formation? This question raises some non-trivial issues, as compounding apparently involves more structural positions and operations than the Hebrew root-and-pattern system. While some theoretical points will be discussed together with the empirical facts in the next section, this section explores an issue of some broader significance: how a word-level phase, henceforth called xP phase for convenience, is to be properly formulated, an issue of some broader significance. More specifically, it is about whether or not roots are allowed to occupy a position other than the complement, such as a modifier position. This issue may actually be separated into two parts: whether a modifier is allowed in an xP phase, and whether a root is allowed in that position.

Beginning with the first part of the issue, we note that Arad's formulation of an xP phase in (2) consists of only two positions: the head and its complement. The modifier, including both the specifier and adjunct, which is an acknowledged constituent in other structures, is notably missing in this framework. In fact, not

---

10. An anonymous reviewer agrees on the validity of Table 1, but suggests that further statistical evidence is needed to make the 'more/less' description more precise. We thank the reviewer for the suggestion, but hold that the 'more/less' tendencies are nonetheless adequate for the purpose of our current discussion, since they are meant just as diagnostics in revealing the properties of the different types of compound.



only Arad, but many other active researchers in DM (for instance, Embick 2004, 2010; Lopez 2015 among others) have chosen not to include such a modifier in their discussion of xP phases. The reason may be that they deal mostly with affixation processes, where the only admissible relation seems to be that of the head-complement, on a par with the root-and-pattern system that we have examined.

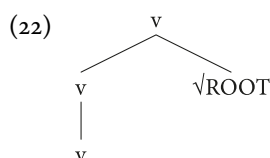
Compounding cases, however, seem to suggest that the modifier position has a vital role to play in our understanding of the processes in word-formation. In the attributive class (see Section 2.2), the non-head does not behave either like the head or its complement. Instead, it aligns with Chomsky's (2001) definition of the specifier. In other words, by admitting the non-head into the specifier position, a better formulation of xP phases could be achieved, since it takes on the same form as the more standard phases in the literature. In fact, it is in line with Chomsky's (2001) formulation of phases, where it is proposed that a phase consists of a head (H), its complement, and an *edge*, which refers to a specifier or an adjunct to H or HP. It would be a conceptually desirable move if we could achieve a uniform formulation of phase in both syntax and word-formation; otherwise the effort to eliminate the generative lexicon is rendered vacuous. Indeed, it is a case of 'conceptual necessities' (Chomsky 1993 and subsequent works) in which there is only one notion of locality for both word-formation and syntax unless forced by empirical considerations.

The second part of the issue, whether a root is allowed in the modifier position, is even more controversial. The prevalent view in the literature is that roots typically occupy the complement position where they are supposed to be categorized. Some previous studies even argue that roots are confined to the complement position (see, for instance, Harley 2009; Borer 2013. Such is also the view implied in Arad's work, as well as in Marantz 2008; Harley 2012, among others). This view seems to be supported by some languages such as English, as they frequently carry functional markings such as participial affixes, as shown in Section 2.2.

Data from Chinese compounds, however, suggest that roots may be inserted in the modifier position in the same way as categorized elements. As shown in Section 2.2, the categorial status of the modifier in Chinese attributives is known to be ambiguous and is hard to identify either as a noun or as an adjective. The modification of the nominal element by a verb-like morpheme without any morphological change, as in (12), suggests that the morpheme involved is category-neutral. Moreover, the presence of a nonsense syllable in the modifier position, as in (13), and the modifier as unique morphemes, as in (14), further suggest that they are mere placeholders with no semantic interpretation of their own. These peculiar properties, which are exhibited in (12)–(14), can easily be taken care of if we treat them as roots devoid of any grammatical features. It is noteworthy that

Chinese attributives have similar properties to the following English compounds where modifiers can also be identified as roots and are therefore incapable of morphological changes: *carefree*, *crybaby*, *diehard*, *pushpin*, *racehorse*, *scarecrow*, *walkathon*, among many others.

Having a root in the modifier position does not violate the spirit of the standard Merge theory. Assuming Chomsky's (2004) distinction between Set Merge, which creates an unordered set, and Pair Merge which creates an ordered pair via adjunction by which a phrase is adjoined to another phrase which projects the label, it may be argued that the root, which never projects the label due to its lack of any grammatical feature, may be introduced into a structure via Pair Merge. This in turn means that it can only function as an adjunct. On the other hand, a root operated on by Set Merge can only be the complement to a head that projects the label. This is essentially the view expressed in Alexiadou and Lohndal (2017: 219–220) who propose that roots may be assumed to interact with categorizers in a similar way that adjuncts interact with the non-adjoined, projecting part of the structure, as shown in the following diagram (adapted from their (39)). Thus, strictly speaking, roots do not have to be categorized. This would indicate that there are cases where roots can survive a derivation without being categorized.



We believe that Alexiadou and Lohndal (2017) are on the right track, since any departure from the move requires extra apparatus to rule out the structure (22), which allows for the existence of an uncategorized root.

Cross-linguistic evidence also seems to support the root-as-modifier treatment. De Belder (2017) argues, for instance, that there are two types of primary compounding in Dutch, with the first invariably containing a root that is nominalized by means of a nominal functional material as its non-head, whereas the non-head in the second type is demonstrably a bare root. These two types are shown in (23), where LP is short for linking phoneme.<sup>11</sup>

- (23) a. The first type with nominal non-head:
- |              |                 |              |              |
|--------------|-----------------|--------------|--------------|
| varken-s-hok | bakker-s-winkel | kat-en-luik  | peer-en-boom |
| pig-LP-pen   | baker-LP-store  | cat-LP-panel | pear-LP-tree |
| 'pig's pen'  | 'bakery'        | 'cat door'   | 'pear tree'  |

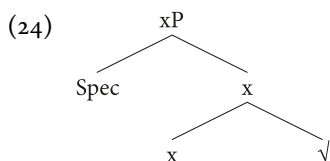
11. De Belder (2017: 141–142), her (4) and (5) respectively.

- b. The second type with bare root non-head:

|              |                |                    |             |
|--------------|----------------|--------------------|-------------|
| kleer-kast   | speur-hond     | snel-trein         | achter-deur |
| cloth-closet | track-dog      | fast-train         | back-door   |
| 'wardrobe'   | 'tracking dog' | 'high-speed train' | 'back door' |

In each instance of (23a), the non-head is followed by an LP, which has long been assumed to instantiate a piece of nominal inflection. By contrast, the non-head in (23b) is directly adjacent to the head, without any intervening material. De Belder (2017) demonstrates that it is not the case that the non-head in the second type is merged with a null categorial head, since the appearance of an intervening overt categorizing affix renders the formation ungrammatical. Given the assumption that null categorial heads and overt categorial heads are only distinguished with regard to the Phonological Form, she concludes that the second type of compound does not contain an intervening null categorial head and that its non-head is not morphologically complex. Moreover, many roots, which serve as the non-head of a compound, are not licit in *bona fide* nominal positions. They occur commonly in both nominal and verbal contexts, suggesting that they need not to be interpreted nominally. Based on such evidence, De Belder concludes that the type illustrated in (23b) contains a bare root as their non-head.

Based on both the conceptual and empirical reasons presented above, we conclude that the version of xP phases in Arad's accounts, attributed to Marantz (2000), needs to be revised so as to include a modifier position, either for a specifier or an adjunct. This is structurally the edge in Chomsky's definition of phase. Such a conception of the xP phase is schematized as (24), according to which root may occupy either the complement position or the modifier position:



In (24), x is the categorizing morpheme, which serves as the head. In the first Merge, it takes a root ( $\sqrt{\phantom{x}}$ ) as its complement, thus forming a more fundamental relation. In the second Merge, the structure assembled by the first Merge merges with another morpheme, possibly a root, forming what is traditionally known as a head-specifier relation.<sup>12</sup>

12. Note that this conception differs from that of De Belder (2017), who proposes that the root primary compounds are not built in syntax, but in PF instead, essentially as a fission process by which the vocabulary item of the root is inserted into the non-head position. We reject such an analysis, but shall restrain from a detailed discussion due to spatial constraints.

Needless to say, such a formulation of a phase configuration is conceptually more satisfactory, for it provides a unified structure with other proposed phases such as CP and  $\nu$ P, and possibly DP as well. Empirically, it accommodates a wider range of facts, especially with respect to attributive compounds, a point that we will return to shortly.

To sum up: we have now pieced together a set of theoretical apparatuses for dealing with word-formation in Chinese, primarily consisting of the following assumptions:

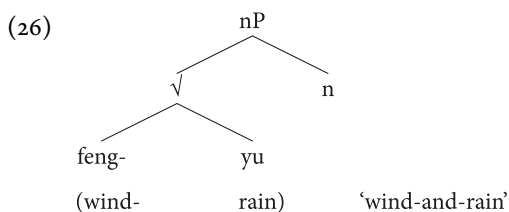
- (25) a. Roots are devoid of any grammatical features, in particular, categorial features. They are interpreted, phonologically and semantically, at a post-syntactic stage, known respectively as PF and LF.
- b. Grammatical features are provided by functional heads. Since the Merge of each head constitutes a phase, which is a closed domain for interpretation, a distinction between word-formation from roots and word-formation from existing words is expected. The former is associated with more idiosyncratic properties – traditionally known as lexical – whereas the latter manifests the more regular properties typically associated with phrases.
- c. Pair Merge may target a root to an adjunct position in the form of the structure in (24), in the same way that it does to a categorized unit.

#### 4. The structure of Chinese compounds

In Section 2, we introduced some basic facts about Chinese compounds. We endeavored to show that different types of Chinese compound exhibit different degrees of syntactic properties with coordinates and attributives sharing more lexical properties while resultatives and subordinates share more phrasal properties. In Section 3, we proposed a modification of the definition of the xP phase, which differs from the one in Arad's account in that it includes a modifier position, parallel to the more standard phases. With the preliminary work in place, we are now in a position to develop an account of the structure for each of the type, in an effort to discover what is behind these different properties.

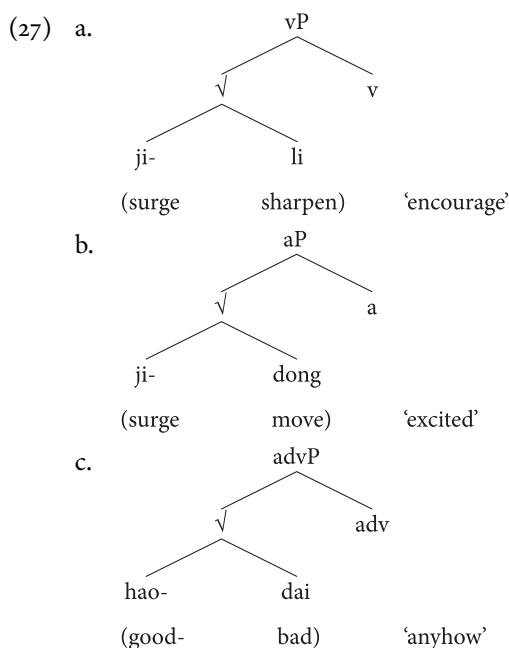
##### 4.1 Coordinate compounds

We assume that Chinese coordinates involve the concatenation of one root with another, which results in a flat structure with neither root dominating the other. This assumption suggests that the configuration of the coordinate compounds should take the following form:



In (26) the roots first combine to form a complex root, which then merges with a category-assigning morpheme. Other instances, such as *feng-shui* (wind-water) 'geomancy', *feng-yue* (wind-moon) 'romance', can be similarly analyzed.

Coordinate compounds can also be verbal, adjectival, or adverbial in terms of their lexical category, as shown respectively in (27):



Examples (26)–(27) epitomize the inner domain in Marantz's (2001) sense, where a root (albeit a complex one) merges with the first category-defining head. Its idiosyncratic properties follow from such a conception in a straightforward manner: the word is categorially unpredictable because it is a scenario in which a root is categorized by the category-assigning head, whatever it is; it is semantically non-compositional precisely because it involves an inner domain; it is also non-interruptible because it is in a phase closed off from outside intervention.

The plausibility of the proposed account is supported by the fact that coordinates other than dissyllabic compounds exhibit the same properties. The following instances are coordinates with three roots. In the first place, they are categorially

unpredictable. In (28a), three adjective-like morpheme sequences, *gao-fu-shuai* ‘tall-rich-handsome’ and *bai-fu-mei* ‘fair-rich-beautiful’, are used as nouns, denoting a particular type of man or woman as indicated in the gloss; in (28b), the adjective-like sequence *gao-da-shang* ‘superior-lofty-classy’ keeps the adjectival feature, shown by the fact that it may co-occur with *hen* ‘very’, a standard adverb that modifies an adjective. It is semantically non-compositional, and inseparable by any foreign material.

- (28) a. *gao-fu-shuai*      *zhi qu*      *bai-fu-mei*  
 tall-rich-handsome only marry white-rich-beautiful  
 ‘tall, rich and handsome men only marry fair-skinned, rich and beautiful women.’
- b. *tade fayan*   *hen*   *gao-da-shang*  
 his speech very superior-lofty-classy  
 ‘his speech is very lofty.’

The discussion in this section thus suggests that coordinates seem to be generated when two roots are concatenated, which is in turn categorized by a category-assigning head. Moreover, the coordinates of more than two syllables seem to share the same properties as the dissyllabic ones, a piece of supporting evidence for the proposed structure.

## 4.2 Attributive compounds

Based on what has been discussed so far, attributive compounds in Chinese involve a modifier-head relation between the two overt constituents. Such a structure is represented by (29).

- (29)
- |       |         |           |                  |
|-------|---------|-----------|------------------|
|       |         | nP        |                  |
|       |         | └───┬───┘ |                  |
|       | √       |           | n                |
|       |         |           |                  |
| hong  | qi      |           | n                |
| (red- | flag)   |           | ‘red flag’       |
|       |         |           |                  |
| hong  | ren     |           |                  |
| (red  | person) |           | ‘popular person’ |

(29) shows that attributive compounds in Chinese are derived in two steps: first, one of the roots (*qi* in *hongqi* ‘red flag’ or *ren* in *hongren* ‘popular person’) merges with a phonetically empty *n*, the nominal category-assigning morpheme.<sup>13</sup>

13. The structure so formed is notated as *n'* in the more traditional X-bar theory.

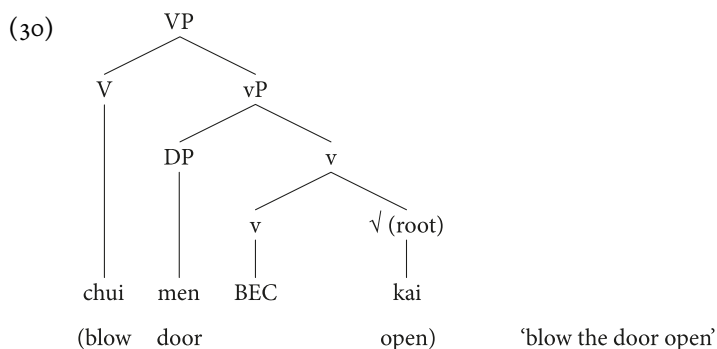
The relation is more fundamental since it is the first Merge that forms a head-complement relation. Second, the combination of the root and the *n* head then merges with a root which takes the specifier position, serving as its modifier. Note that, according to our revised definition in (24), these merging operations take place in the same phase, where the modifier is conceived as its edge.

Such an analysis effectively captures the major properties of attributive compounds. In the first place, its apparent endocentricity is predicted, since it is the second overt morpheme that is merged with the category-assigning morpheme; the modifier, by contrast, is incorporated in the structure via the second Merge. In the second place, its semantic idiosyncrasy is also captured, given the fact that the modifier is a root that is in turn at the lowest level of semantic interpretability. Finally, its internal cohesion comes as no surprise either, since the root modifier is by definition featureless, and is thus unable to undergo any syntactic operations.

### 4.3 Resultative compounds

This subsection focuses on how to capture the major properties of resultative compounds – high endocentricity, semantic transparency, and a low degree of interruptability – as sketched in Section 2.3. We also wish to deal with two other important facts mentioned in that section. One is the parallelism between resultatives and SVCs, which we take as an indication that the two constructions are generated from the same underlying structure, and that resultative compounds should be derived by verb movement from an underlying structure that is similar to that of SVCs. The other fact is that V2 is ambiguous in its categorial status and cannot be identified as either a verb or an adjective.

Taking all these facts into consideration, Cheng and Yang (2016) propose that resultative compounds have the schematic configuration shown in (30), where BEC(ome) stands for a null functional head denoting a change of state.



Accordingly, both resultative compounds and SVCs have the underlying structure of a complex predicate (VP). V denotes the action and licenses a vP which denotes

the result. The vP, in turn, consists of three subparts (cf. Embick 2004; Huang 2006; Xiong 2015, and references therein): (i) BEC; (ii) a DP at its specifier position, generally interpreted as the theme; and, importantly, (iii) a root that is directly merged with BEC. At the post-syntactic stage, the [v √] combination undergoes the morphological process of compounding, by which it is fused with the action-denoting V, yielding the surface form of resultative compounds. In other words, resultative compounding involves a post-syntactic operation that adjusts a syntactic structure and forms a lexical item.

Major advantages of the proposed account include the following: firstly, it is a syntactic model that derives resultative compounds and SVCs simultaneously; secondly, it captures the semantic interpretation of the construction, to wit, an event which involves a change of state as commonly assumed; thirdly, it predicts the morpho-syntactic behaviors of the resulting-denoting element by identifying it as a root which is by definition devoid of grammatical features. By contrast, most existing accounts take the element as a verb, notated mistakenly as V2. Such accounts fail to capture the ambiguous status that the morpheme manifests in this environment; and finally, the idea of compounding as a post-syntactic operation is compatible with the generally-accepted view that resultative compounding is primarily motivated by the morpho-phonological requirement which forces words in Modern Chinese to be disyllabic.

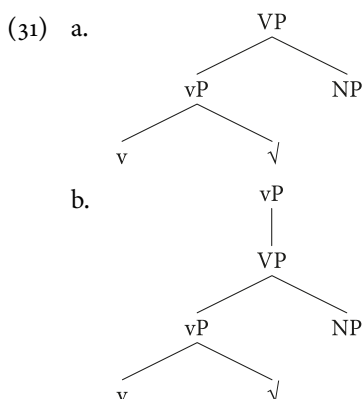
The present account thus captures the major properties of resultative compounds. By assuming that these compounds are derived by the merging of a verb and its licensed complement which is in turn composed of a null head (BEC) and a root, it follows that the structure is determined by its head – V1, and is thus endocentric; semantic transparency is also expected given the head-complement relation between the two elements; its strong internal cohesion receives a natural explanation as well since there is a minimal link between V and v, which prevents the intervention of a more distant head. If this line of reasoning is on the right track, such compounds provide a strong piece of evidence for the syntactic approach to word-formation and a grave threat to the lexicalist approach that is apparently unable to generate the SVC structure, which cannot be built in the pre-syntactic lexicon in any conceivable way.

#### 4.4 Subordinate compounds

In this subsection, we examine the last major type of compound – subordinate compounds. As was pointed out in subsection 2.4 above, they bear a close resemblance to phrases with the highest degree of endocentricity and semantic compositionality but the least internal cohesion.



We believe that subordinate compounds illustrate the DM thesis of morphology as syntax, in that they act as a verb phrase turned back to a lexical item through so-called *lexicalization* (see the discussion in Section 2.4), a process by which frequently co-occurring sequences gradually become words in the traditional sense. In more formal terms, we believe that the majority of subordinate compounds are derived with two steps (cf. Footnote 7). The first is the making of a VP, and the second a reanalysis of it to vP, shown respectively by the following configurations:<sup>14</sup>

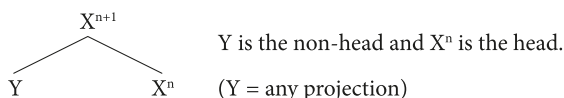


The proposed account is a natural explanation for the unique properties of subordinate compounds. According to the assumptions laid out in Section 3, VP is an outer domain where the root is no longer visible. It is thus morpho-syntactically more regular and semantically more transparent, which in turn gives rise to the categorial endocentricity and semantic compositionality that are manifested in the structure. Moreover, given the locality constraints, syntactic operations cannot backtrack previous derivations. In other words, the phrasal properties of such compounds can never be canceled and will surface whenever they are placed in a proper environment or needed to perform discursal or pragmatic functions. This is the reason that VO compounds frequently behave as phrases, like their ability to be used as a so-called separable word.

The proposed account is additionally supported by the stress properties. Duanmu (1990) famously suggests that stress is assigned in accordance with the ‘Non-head Stress (NHS)’ rule, by which a syntactic non-head receives stress in a spell-out domain, schematized as follows (Duanmu 1990: 174):

14. Unlike Zhang (2007), we believe that (31) is the structure into which pseudo-VO compounds are reanalyzed. She views instances such as *shui jiao* ‘sleep sleep’ as parallel combination strings and claims that they behave like compounds in one context and phrases in another. Yet the fact that *jiao* can take a classifier suggests that it is nominal and is thus not a parallel of *shui*.

(32) In



The NHS rule succeeds in subsuming a wide range of facts in different types of construction in Chinese, at both the traditional lexical and phrasal levels. At the lexical level, there are instances from nominal compounds such as *XIAN-nai* ‘fresh milk’, *YOU chao-fan* ‘cooked food with oil’ (capital letters are used to indicate the stress placement); the verbal: *HONG-shao* ‘braise in soy sauce’, *SHOU-xie* ‘write with hand’, as well as the adjectival compounds: *BING-leng* ‘ice-cold’, *SHEN-lan* ‘deep blue’. At phrase level, there are verbal instances such as *chao-FAN* ‘cook rice’, *he-SHUI* ‘drink water’, *he XIAN niu-nai* ‘drink fresh milk’, as well as prepositional phrases such as *zai WAI-mian* ‘at out-side’.

Subsequent studies, however, reveal some anomalies with the NHS. For the purpose of the present discussion, we will focus on the homophone stress patterns, illustrated in (33)–(35):

(33) kao ROU

roast meat

a. to roast meat

b. roasted meat

(34) chao FAN

stir-fry rice

a. to stir-fry rice

b. stir-fried rice

(35) zheng DAN

steam egg

a. to steam an egg

b. a steamed egg

As shown in the translation, the sequences in (33)–(35) are ambiguous between two interpretations, either as verb phrases, as indicated in (a), or as nominal compounds, as indicated in (b). This in turn means that they have different heads in the reversed order: the former is the verb at the left, and the latter the noun at the right. The stress placement in these pairs is consistently on the last morpheme, however, as illustrated by the capital letters, which seems to contradict the NHS theory.

Hai Li (2016), however, shows that the anomaly is only apparent and that it disappears if derivation by phase is assumed for these structures. They all involve the formation of a VP as the first step, as in (31a). In this process, stress is assigned to the object, in accordance with the NHS. Since subsequent derivation cannot

retract the computation, the stress gets stuck with the object, even if it does not undergo the reanalysis of the VP into a vP, as in (31b), or a further process of nominalization which yields the (b) cases in (33)–(35).

These facts concerning the stress pattern are significant. On the one hand, they give strong evidence that the genesis of subordinate compounds is VP, composed of a verb and its complement so that the latter receives the stress. On the other hand, they indicate that word-formation is derived by phase, the effects of which cannot be canceled in subsequent derivations.

To summarize the discussion thus far, subordinate compounds in Chinese are derived from an existing phrase via the morphological process of lexicalization, as illustrated in (31). This origin gives rise to its phrase-like properties, such as categorial endocentricity, semantic transparency, and separability. In particular, it is also the cause for its use both as a word and a phrase. Moreover, if Hai Li (2016) is right, this origin also regulates the stress placement in subordinate compounds.

Our analysis of subordinate compounds, if on the right track, adds evidence to the Single Engine Hypothesis that words and phrases are assembled by essentially the same generative system (the syntax). It also confirms the role of locality constraint in word-formation, which bars the outcome of its derivation from being canceled once it is transferred to LF and PF.

#### 4.5 The continuum of structural properties revisited

Now let us take a look back at the major types of compound under discussion in the present study. Recall the claim in Section 2.1 that there is a continuum from coordinates, the most lexical in character, to subordinates, the most phrasal, with attributive and resultative compounds being located somewhere in the middle. Based on the previous discussion, we believe that we can account for the continuum as follows: coordinate compounds are the most lexical because they are derived from the merger of two roots, as a case of word-formation from roots; attributive compounds are less lexical because they involve the merging of a word and a root; a resultative compound is more phrasal because it is the merger of two words, a *prima facie* instance of word-formation from words; finally, subordinate compounds are the most phrasal, because they are indeed derived from the lexicalization of existing phrases. (36) below is a schematized illustration of the proposed account.

|      |                 |             |             |             |
|------|-----------------|-------------|-------------|-------------|
| (36) | WORDS → PHRASES |             |             |             |
|      | coordinate      | attributive | resultative | subordinate |
|      | root + root     | root + word | word + word | phrase      |

## 5. More on phases

This section continues our discussion on the effect of phases in Chinese compounds with the dual purpose of demonstrating that (i) phase is indeed crucially involved in Chinese word-formation, and (ii) the revised version of phase as in (24) is indeed superior to that in existing literature such as (2). We will try to achieve this by taking a closer look at attributive constructions. In addition to the direct combination of a modifier with a noun, henceforth notated as '[M N]', there is also a type where the combination between the modifier and noun is mediated by *de*, the modifier marker, notated here as '[M *de* N]'. It has long been noted that systematic differences exist between two structures, summarized as follows (see Fan 1958; Huang 1984; Dai 1992; Duanmu 1998 and references cited therein).

Firstly, conjunction reduction applies only to [M *de* N], but never to [M N], illustrated below.

- (37) a. [bai *de* yi-fu] he [bai *de* xiezi]  
 white *DE* clothes and white *DE* shoe  
 'white clothes and white shoes'  
 b. bai *DE* [yi-fu he xiezi]  
 white *DE* clothes and shoe  
 'white [clothes and shoes]'
- (38) a. [bai cai] he [bai rou]  
 white vegetable and white meat  
 'cabbage and plain boiled meat'  
 b. \*bai [cai he rou]  
 white vegetable and meat  
 'white [vegetable and meat]'

(37a) involves two [M *de* N] structures, and conjunction reduction may apply to give (37b). In contrast, (38a) involves two [M N] structures and conjunction reduction is impossible, as shown in (38b).

Secondly, semantic composition is a necessary feature of [M *de* N], but not of [M N]. Thus, (39a) means 'a board that is black'. As such, it cannot be modified by an additional adjective *bai* 'white' which contradicts the original meaning, as shown in (39b). In contrast, *hei-ban* in (40a) is a compound, meaning 'black-board'. Adding *bai* 'white' (with or without the particle *de*) is possible, even though *bai* 'white' contradicts *hei* 'black'.

- (39) a. hei *DE* ban  
 black *DE* board  
 'white black board'

- b. \*bai hei DE ban  
white black DE board
- (40) a. hei-ban  
black-board  
'blackboard'
- b. bai DE hei-ban  
white DE black-board  
'white blackboard'
- c. bai hei-ban  
white black-board  
'white blackboard'

Thirdly, adverbial modification applies to M in [M *de* N], but not [M N]. In the instances to follow, an adverb, typically an adverb of degree such as *hen* 'very', is allowed in [M *de* N], but banned in [M N], as first noted in Fan (1958: 214).

- (41) a. xin DE shu  
new DE book  
'a new book'
- b. hen xin DE shu  
very new DE book  
'a very new book'
- c. geng xin DE shu  
more new DE book  
'a newer book'
- d. zui xin DE shu  
most new DE book  
'the newest book'
- (42) a. xin shu  
new book  
'a new book'
- b. \*hen xin shu  
very new book  
'a very new book'
- c. \*geng xin shu  
more new book  
'a newer book'
- d. \*zui xin shu  
most new book  
'the newest book'

Fourthly, there is a contrast between the two structures in terms of substitution by a phrase. Fan (1958: 214) notes that N in [M *de* N] can be substituted by [X N] where X is a numeral-classifier unit or a demonstrative. However, N in [M N] cannot be substituted in this way. This contrast is shown in the following examples.

- (43) a. xin DE [san ben shu]  
new DE three copy book  
'three books that are new'  
b. xin DE [nei ben shu]  
new DE that copy book  
'that book which is new'
- (44) a. \*xin [san ben shu]  
new three copy book  
b. \*xin [nei ben shu]  
new that copy book

Facts presented above show that [M *de* N] and [M N] structures behave differently with respect to conjunction reduction, semantic compositionality, adverbial modification, and substitution by a phrase. Systematic differences between the two are evidenced by other facts, such as productivity and insertion, a topic that is beyond the scope of the present discussion. For lexicalists, these facts are taken as suggesting that [M *de* N] is always a phrase, whereas [M N] is most likely to be a word (cf. Duanmu 1998).

We believe there is no need to accept the view that phrases and words are fundamentally different structures or the view that grammar incorporates a Lexical Integrity Principle (LIP), which forbids syntactic rules from applying to the internal structure of words. Instead, the contrast between [M *de* N] and [M N] can be derived from the xP phase theory as formulated in (24).

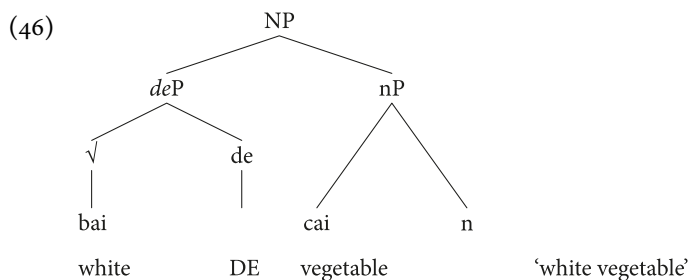
As an illustration, let us observe the way in which the structures of *bai-cai* 'cabbage' [M N] and *bai de cai* 'white vegetable' [M *de* N] are generated. In line with the analysis in 2.4, *bai-cai* has the structure given in (45), where the little n takes *cai* as its complement and *bai* as its modifier.

- (45)
- 
- ```

graph TD
    nP[nP] --- v1[√]
    nP --- n1[n]
    v1 --- bai[bai]
    bai --- White[White]
    n1 --- v2[√]
    n1 --- n2[n]
    v2 --- cai[cai]
    cai --- vegetable[vegetable]
    n2 --- n3[n]
    n3 --- cabbage["'cabbage'"]
  
```

The structure in (45) constitutes a phase, which includes a head (little *n*), its complement –  $\sqrt{\text{cai}}$  ‘vegetable’, and a specifier –  $\sqrt{\text{bai}}$  ‘white’.

The same cannot hold for  $[\text{M } de \text{ N}]$  structures, however. It has long been accepted that *de* is a functional category – a particle of structure (*jiegou zhuci*). As a matter of fact, Simpson (2001), Tsai (2003), Xiong (2005) and many others have more recently argued that *de* is a determiner, though a bleached and enclitic one. These theories point to the fact that *de* is a head projecting its properties, which in turn means that the  $[\text{bai } de]$  structure forms a phase from its lexical subarray (numeration), independent of *cai*. Their structure is shown below:



It is noteworthy that the more recent *de*-as-a-determiner hypothesis accords well with the present account since DP is widely believed to be a higher-order phase. However, we will refrain from pursuing the point further here. Suffice it to say that, so long as *de* is a functional category, it is not in the same phase as N in the  $[\text{M } de \text{ N}]$  structure.

To conclude: the present account interprets the systematic differences between  $[\text{M } de \text{ N}]$  and  $[\text{M } \text{N}]$  essentially as a locality effect, namely,  $[\text{M } \text{N}]$  is a closed domain where the structure is transferred to LF and PF. As such, its semantics tends to be idiosyncratic since they involve negotiating the root meaning, giving rise to the non-compositional effect illustrated in (39b); its structure is closed to any further syntactic computation, so that conjunction reduction, adverbial modification, substitution by a phrase, and other syntactic operations are disallowed. As such, the  $[\text{M } \text{N}]$  structure tends to be perceived as lexical. On the other hand, the  $[\text{M } de \text{ N}]$  structure can only be phrasal, since it contains two phases, one of which headed by *de*, assembled with different subarrays (numerations).

This explains why  $[\text{M } de \text{ N}]$  structure is more compositional in semantics and is able to undergo syntactic operations such as reduction, adverbial modification, or movement. Accordingly, the unnatural distinction between phrases and words, together with the LIP, may no longer be viewed as theoretical primitives but are rather derivable from deeper principles. Moreover, it also lends support to the revised formulation of the phase in (24) for its increased empirical coverage.

## 6. Conclusion

This paper provides a syntactic account of the continuum from word-like properties to phrase-like properties in Chinese compounds, it shows that the coordinate type is the most lexical and the subordinate type is at the other extreme, while the attributive and the resultative are situated somewhere in the middle (see Table 1 in Section 2.5). It shows that the continuum is brought about by syntactic operations in conjunction with locality constraints – more specifically, lexical properties come from the inner domain with roots as input to the derivation, and phrasal properties from the outer domain with words as input. Coordinates and attributives are relatively word-like because they instantiate derivation from roots, whereas resultatives and subordinates are derived from existing words.

If correct, the present account lends support to the fine-grained approach to the Hebrew root-and-pattern system advocated by Arad (2003, 2005). It suggests that despite apparent differences, Chinese and Hebrew actually form words with essentially the same operations and processes, which are subject to the same locality constraints. In fact, accumulating evidence seems to suggest that similarities between these two languages extend to other languages; in particular, the distinction between word-formation from roots and words is supported in many subsequent studies (see, among others, Brice 2017 and references therein).

The present account bears witness to the need to re-examine the traditional view that linguistic variation is located in the lexicon (cf. Borer 1984), where the lexicon is conceived as including all sorts of pre-syntactic rules for word-formation. Words in Chinese and Hebrew are of course vastly different, and there seems little hope that they could be treated in a unified way. However, as can be seen in the paper, the basic rules for building words in the two languages are essentially the same, as are locality constraints on their application. Furthermore, the rules and constraints are syntactic, with consequences well beyond the lexical level. If this approach is on the right track, there is no need to posit pre-syntactic rules and processes that are distinct from, and independent of, syntactic ones.

As for the question of what makes Chinese different from Hebrew in terms of word-formation, we believe that the primary source is to be found in morphemes, which come in two varieties. On the one hand, roots, the substantive morphemes, may be taken as a source of variation, as is widely acknowledged. As seen in the foregoing discussion, Chinese roots are mostly CV sequences, perfectly pronounceable on a par with words. In fact, from a historical point of view, in Modern Chinese many roots have been recruited from words in Ancient Chinese, at which time words and roots were both monosyllabic (cf. Feng 1997). In this respect, Chinese roots stand in stark contrast to Hebrew roots, which consist of segmental consonants and are manifestly different from words. The two languages



also differ in the number of roots available, with Chinese seemingly possessing a larger number of roots than Hebrew. Since each Hebrew root is responsible for building a large number of words, its meaning is naturally variable, giving rise to what is dubbed by Arad (2003) as ‘multiple contextualized meaning’. By contrast, Chinese roots seem to be semantically more concrete despite their capacity for multiple interpretations as well (see in particular Section 2). On the other hand, greater difference seems to exist in the inventory of functional morphemes. A large number of features, which are morphologically realized in Hebrew, are absent in Chinese, including nominal features such as case and agreement (person, gender, and number) as well as verbal features such as tense, voice, and mood. In addition, Hebrew is able to express causative, transitive, reflexive, reciprocal, passive, and iterative meanings through the prefixation or infixation of consonantal elements, as well as an array of phonological alterations (see Rubin 2010 and references therein for a detailed discussion). As it has been an analytic language throughout its history, Chinese does not possess such morphological means. As a result, many meanings or functions are expressed syntactically or without morphological markings. This is especially true of the cases that involve dissyllabic compounds. As can be seen in the previous discussion, these compounds are not overtly marked for any grammatical information. Judging by evidence from contemporary studies, we might safely conclude that the property of morphemes ultimately makes a root-and-pattern system impossible in Chinese, whereas the adoption of compounding as its primary means of word-formation is extremely difficult in Hebrew. In other words, the locus of linguistic variation is the morpheme. The rules that combine them into words and other larger units, as well as the phase-based nature of the rules, are fundamentally the same.

## References

- Aikhenvald, Y.A. & Robert, M. W.D. (2006). *Grammars in Contact – A Cross-Linguistic Typology*. Oxford: Oxford University Press.
- Alexiadou, A. & Terje, L. (2017). The structural configurations of root categorization. In L. Bauke & A. Blümel (Eds.), *Labels and Roots*: 203–232. Berlin: De Gruyter.  
<https://doi.org/10.1515/9781501502118-009>
- Arad, M. (2003). Locality constraints on the interpretation of roots: the case of Hebrew denominal verbs. *Natural Language and Linguistic Theory* 21: 737–778.  
<https://doi.org/10.1023/A:1025533719905>
- Arad, M. (2005). *Roots and Patterns: Hebrew Morphosyntax*. Dordrecht: Springer.
- Baker, M. (1989). Object sharing in serial verb constructions. *Linguistic Inquiry* 20: 513–553.
- Bisetto, A. & Sergio, S. (2005). The classification of compounds. *Lingue e Linguaggio*, 4 (2): 319–332.
- Bloomfield, L. (1933). *Language*. New York: Holt.

- Borer, H. (1984). *Parametric Syntax*. Dordrecht: Foris Publications.  
<https://doi.org/10.1515/9783110808506>
- Borer, H. (2013). *Taking Form*. Oxford: Oxford University Press.
- Brice, H. (2017). The root and word distinction: an experimental study of Hebrew denominal verbs. *Morphology* 27: 159–177. <https://doi.org/10.1007/s11525-016-9297-0>
- Cao, C. (2016). Xiandai Hanyu Pianyi Fuci Tanjiu [On Partially-directed Compound Words in Modern Chinese]. MA Thesis, Shanghai Normal University.
- Chao, Y. (1968). *A Grammar of Spoken Chinese*. Berkeley, California: University of California Press.
- Cheng, G. & Yang, D. (2016). Hanyu dongjieshi fuheci de yuxu ji xiangguan wenti [Word order in Chinese VV compounds and related issues]. *Zhongguo Yuwen* 374 (5): 553–567.
- Cheng, G. & Zhou, G. (2015). Verb-object compounding in Chinese: A Distributed Morphology approach. *Foreign Language Teaching and Research* 47: 163–175.
- Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale and S. J. Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*: 1–52.
- Chomsky, N. (1998). *Minimalist Inquiries: The Framework*. Ms: MIT.
- Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Eds.), *Ken Hale: A Life in Language* (pp. 1–52). Cambridge: MIT Press.
- Chomsky, N. (2004). Beyond explanatory adequacy. In A. Belletti (Ed.), *Structures and Beyond* (pp. 104–131). Oxford: Oxford University Press.
- Collins, C. (1994). Topics in Ewe Syntax. Ph. D dissertation, Cambridge, MA: MIT Press.
- Collins, C. (2002). Multiple verb movement in #Hoan. *Linguistic Inquiry* 33: 1–29.  
<https://doi.org/10.1162/002438902317382161>
- Dai, X. (1992). Chinese Morphology and its Interface with the Syntax. Ph.D. dissertation, Columbus: Ohio State University.
- De Belder, M. (2017). The root and nothing but the root: primary compounds in Dutch. *Syntax*: 1–32.
- Dong, X. (2009). Hanyu de jufa yanbian yu cihuihua [Syntactic change and lexicalization in Chinese]. *Zhongguo Yuwen* 332 (5): 399–409.
- Duanmu, S. (1990). A formal study of syllable, tone, stress and domain in Chinese languages. Ph.D. dissertation. Cambridge: Massachusetts Institution of Technology.
- Duanmu, S. (1998). Wordhood in Chinese. In J. L. Packard (Ed.), *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese* (pp. 135–196). Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110809084.135>
- Embick, D. (2004). On the structure of resultative participles in English. *Linguistic Inquiry* 35: 355–392. <https://doi.org/10.1162/0024389041402634>
- Embick, D. (2010). *Localism versus Globalism in Morphology and Phonology*. Cambridge: MIT Press. <https://doi.org/10.7551/mitpress/9780262014229.001.0001>
- Fan, J. (1958). Xing-ming zuhe jian ‘de’ zi de yufa zuoyong [The grammatical function of *de* in adjective-noun constructions], *Zhongguo Yuwen* (5): 213–217.
- Feng, S. (1997). Hanyu de Yunlv, Cifa yu Jufa [Prosody, Morphology and Syntax of Chinese]. Beijing: Peking University Press.
- Harley, H. (2009). The morphology of nominalizations and the syntax of vP. In A. Giannakidou & M. Rathert (Eds.), *Quantification, Definiteness, and Nominalization* (pp. 320–342). Oxford: Oxford University Press.

- Harley, H. (2012). Lexical decomposition in modern generative grammar. In M. Werning, W. Hinzen & E. Machery (Eds.), *The Oxford Handbook of Compositionality* (pp. 328–350). Oxford: Oxford University Press.
- Hopper, P.J. & Traugott, E. C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.
- Huang, C.-T. J. (1984). Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association* 19 (2): 53–78.
- Huang, C.-T. J. (2006). Resultatives and unaccusatives: a parametric view. *Bulletin of the Chinese Linguistic Society of Japan* 253: 1–43. <https://doi.org/10.7131/chuugokugogaku.2006.1>
- Li, H. (2016). Exploration of Syntax-Phonology Interface in Chinese: Perspectives from Phase Theory and Distributed Morphology. Ph.D. dissertation. Luoyang: PLA University of Foreign Languages.
- Li, Y. (1990). On V-V compounds in Chinese. *Natural Language and Linguistic Theory* 8: 177–207. <https://doi.org/10.1007/BF00208523>
- Li, Y. (1995). The thematic hierarchy and causativity. *Natural Language and Linguistic Theory* 13: 255–282. <https://doi.org/10.1007/BF00992783>
- Li, J. (2012). Jiyu yuliaoku de dongbin fuheci shuxing kaocha jiqi cihuihua chengdu fenxi [A corpus-based study of the properties of VO compounds and their lexicalization]. *Xiandai Yuwen* (10): 42–48.
- Lopez, L. (2015). Parallel computation in word-formation. *Linguistic Inquiry* 46 (4): 657–701. [https://doi.org/10.1162/LING\\_a\\_00197](https://doi.org/10.1162/LING_a_00197)
- Lu, Z. (1957). *Hanyu de Goucifa* [Chinese Morphology]. Beijing: Kexue Chubanshe.
- Marantz, A. (2001). Words. Unpublished manuscript. Cambridge: MIT.
- Marantz, A. (2008). Phases and words. In S.-H. Choe (Ed.), *Phases in the Theory of Grammar* (pp. 191–222). Seoul: Dong In Publisher.
- Packard, J. L. (2000). *The Morphology of Chinese: A Linguistic and Cognitive Approach*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486821>
- Rubin, A. D. (2010). *A Brief Introduction to the Semitic Languages*. New Jersey: Gorgias Press.
- Shiao, W. (2015). Resultative verb compounds in Mandarin. In W. S.-Y. Wang & C. Sun (Eds.), *The Oxford Handbook of Chinese Linguistics* (pp. 306–321). Oxford: Oxford University Press.
- Simpson, A. (2001). Definiteness agreement and the Chinese DP. *Language and Linguistics* 2: 125–156.
- Thompson, S. (1973). Resultative verb compounds in Mandarin Chinese: A case for lexical rules. *Language* 49: 361–379. <https://doi.org/10.2307/412459>
- Tsai, W.-T. D. (2003). Three types of existential quantification in Chinese. In A. Li & A. Simpson (Eds.), *Form, Interpretation and Functional Structure: Perspectives from Asian Languages*. London: Curzon/Routledge.
- William, A. (2008). Word order in resultatives. In C. Chang & H. Haynie (Eds.), *Proceedings of WCCFL 26*. Berkeley: University of California.
- Xiong, Z. (2005). Yi “de” wei hexin de DP jegou [A DP structure headed by *de*]. *Dangdai Yuyanxue* 7(2): 148–165.
- Xiong, Z. (2015). Yinghan Zhishi Jufa Lunyuan Jiegou de Duibi Yanjiu [A Contrastive Study of Argument Structures in Causative Construction in English and Chinese]. Shanghai: Shanghai Foreign Languages Education Press.
- Yang, X. (2003). *Yusu Lun* [On Morphemes]. Nanjing: Nanjing University Press.

- Zhang, N. (2007). Root merger in Chinese compounds. *Studia Linguistica* 61(2): 170–184.  
<https://doi.org/10.1111/j.1467-9582.2007.00131.x>
- Zhou, G. & Cheng, G. (2018). How are VN compounds in Chinese derived? *Dangdai Yuyanxue* 20 (1): 61–84.
- Zhou, J. (2005). *Hanyu Cihui Jiegou Lun* [On the Structure of Chinese Words]. Shanghai: Shanghai Cishu Chubanshe.
- Zhou, Y. (2015). A Research of the Subject-Predicate Compounds Words in Contemporary Chinese. MA Thesis. Jinan: Shandong University.

## Address for correspondence

Cheng Gong  
 Department of Linguistics and Translation  
 Zhejiang University  
 No. 866, Yuhangtang Road  
 Hangzhou, Zhejiang 310058  
 P. R. China  
[chenggong@zju.edu.cn](mailto:chenggong@zju.edu.cn)

## Co-author information

Liu Ying  
 Department of Linguistics and Translation  
 Zhejiang University  
[chunhunaxty@hotmail.com](mailto:chunhunaxty@hotmail.com)