Morphological Semi-Phrasemes in Mandarin Chinese: an Atypical Collocational Phenomenon

Étienne Van Tien Nguyen
OLST – Université de Montréal
C.P. 6128, succ. Centre-ville
Montréal (Québec) H3C 3J7 CANADA
vtetienne@yahoo.fr

Abstract

This paper presents an atypical form of “collocation” found in Mandarin Chinese. It aims at showing that in this language, and probably in others with similar features, “collocational phenomena” do not occur only at the syntagmatic level (i.e. within a phrase), as in Western languages such as English or French, but also at the morphological level, within a wordform. This is due to the large number of semantically compositional compound wordforms in Chinese. This paper uses two criteria distinguishing wordforms from phrases to demonstrate that Chinese has semantically compositional compounds: the external modifier criterion and the non-conjunction reduction criterion. It then shows that among these compositional compounds, some are “morphological” semi-phrasemes: the combinatorial properties of one of their components are affected by some phraseologization. This paper also discusses how “morphological” semi-phrasemes should be treated in an Explanatory Combinatorial Dictionary of Chinese.

Keywords


Introduction

In lexicographic studies, the term “collocation” is often used to refer to a phrase [≡ a sequence of at least two wordforms which are syntactically linked] made of lexemes A and B, where the speaker freely selects A for its meaning and is forced to select B from a restricted choice of lexemes to express the meaning ‘C’ or ‘B’\(^1\). For example, in the phrase heavy rain – identified as a collocation in English – rain is freely selected by the speaker to express the meaning ‘rain’, but the choice of heavy to express the meaning ‘intense’ is simply dictated by the language. One cannot say *intense rain in English, or *fat/big rain as is possible in

\(^1\) We arbitrarily chose A as the free element of the phrase: we could have taken B.
French *grosse pluie*. Collocations in English and French have already been described in detail by various scholars, see (Makkai, 1972), (Nunberg, Sag & Wasow, 1994), (Mel’čuk, 1998, 2003). Collocations in Chinese, however, have not been studied extensively. Chinese, like other languages, does indeed have typical collocations such as *jīlì de zhàndòu* be.fierce DE$_1$ battle ‘a fierce battle’. In this phrase, *jīlì* ‘be.fierce’ is a qualifying verb [= QV]$^3$ that modifies *zhàndòu* ‘battle’. It is used here to convey the meaning ‘be.intense’. It cannot be replaced by quasi-synonyms like *xiǒngměng* ‘be.ferocious’ or *qiángbào* ‘be.violent’. So the selection of *jīlì* is imposed by the language. In addition to these typical collocations, Chinese has numerous expressions like *dà-yǔ* be.big-rain lit. ‘big rain’ = ‘heavy rain’. These expressions have the same characteristics as collocations, but they are in fact compound wordforms. In the case of *dà-yǔ*, *-yǔ* is freely selected for its meaning, but to express the meaning ‘be.intense’, *-yǔ* ‘rain’ must be combined with *dà-* ‘be.big’, and cannot be combined with *zhòng-* ‘be.heavy’ as in the corresponding English phrase *heavy rain*, nor by *pàng-* ‘fat’, as in the French equivalent *grosse pluie* lit. ‘fat rain’. Although it is a wordform, *dà-yǔ* shows significant similarities to a collocation. So how does one know that *dà-yǔ* is a wordform and not a phrase? This paper will attempt to provide an answer to this question.

Section 1 shows that Chinese does indeed have compositional compound wordforms [= CC]; Section 2 examines two cases of CCs, such as *dà-yǔ*, that are morphological semi-phrasemes. These will be referred to as “collocational compositional compounds” [= CCC]. Finally this paper addresses the question of how CCCs should be treated in a Chinese encoding dictionary such as an Explanatory and Combinatorial Dictionary [= ECD]$^4$ of Chinese.

1 Compositional compound wordforms in Chinese

1.1 Wordform versus Phrase

Many studies like (Chao, 1968), (Li & Thompson, 1981) and (Yip, 2000) already point out that the Chinese lexicon consists primarily of compounds. But most of the time, what they give as examples of compounds are semantically non-compositional compound wordforms [= NCC], compounds whose meaning cannot be inferred from the meaning of their constituents, as in (1):

---

$^2$ The clitic DE$_1$ always marks the element it cliticizes to (i.e. the element after which it is placed) as a modifier of a noun.

$^3$ A qualifying verb or *xingrongci* in Chinese is a verb that has the same semantic properties as an adjective in such languages as French or English. It has long been debated whether *xingronci* are verbs or adjectives. Following (McCawley, 1992), (Sproat & Shih, 1988, 1991), (Duanmu, 1998), we argue that they are a subclass of verbs, and provide arguments for this hypothesis in (Nguyen, 2006).

$^4$ For a detailed description of what a traditional ECD is, see (Mel’čuk et al., 1984, 1988, 1992, 1999), (Mel’čuk et al., 1995), and (Steele, 1990).
But little has been said about complex signs such as those in (2) and (3), whose meanings are compositional. This might be because they have wrongly been regarded as phrases.

The complex signs in (2) are made of a qualifyi ng verb [= QV] stem and a nominal stem [= N] and those in (3), of two nominal stems. I argue that those in (2) and (3) are compound wordforms and not phrases and that the difference between the complex signs in (1) and those in (2)-(3) mainly lies in their semantic compositionality. The reason why those in (2)-(3) have never been unanimously defined as compound wordforms is that it is very difficult to distinguish a phrase from a wordform in Chinese. This is due to the fact that Chinese does not have morphological markers of composition, since it has very little morphology to speak of. Unlike languages such as German or Russian, Chinese lacks morphological markers that enable one to distinguish a wordform from a phrase. For example, the German wordform *Universitätsgebäude* ‘university+s+building’ is a compound wordform because the -s- marker would not be present in a phrase consisting of the corresponding separate wordforms. Even in the cases of adjective-noun composition in German, where the -s- marker is missing, it is still possible to find morphological markers distinguishing phrases from compounds. For example, the compositional compound *Fremdsprachen* = ‘foreign languages’ can be distinguished from the phrase *fremde Sprachen* = ‘foreign languages’ by the fact that the stem *Fremd-* in the compound wordform remains invariable, whereas the qualifying adjective in German must agree with the noun, which is shown by the ending.

In Chinese, however, one has to use purely syntactic criteria to distinguish a compound wordform from a phrase. To prove that the complex signs in (2)-(3) are compound wordforms and not phrases, I will use two formal criteria: the external modifier criterion and the non-conjunction reduction criterion. These two criteria show that the complex signs in (2)-(3) behave like “syntactic islands,” meaning that their constituents cannot be affected by any syntactic rules.

---

There are more criteria and considerations that help distinguish a wordform from a phrase. We will not be able to talk about all of them, for lack of space. Interested readers can consult (Nguyen, 2006) where these criteria have been discussed in more detail.
1.2 External modifier criterion

The external modifier criterion can be formulated as follows: S₁-S₂ is a compound wordform if S₁ has lost its ability to accept external modifiers. Otherwise, it is a phrase. This criterion has already been proposed in different forms in studies such as (Fan, 1958), (Huang, 1984)⁶ or (Duanmu, 1998).

1.2.1 QV-N expressions

To show that the complex signs in (2) are QV-N compound wordforms, let’s apply the external modifier criterion to one of them, e.g. dà-shù be.big-tree ‘big tree’. As shown in (4a-c), dà- ‘be.big’ within dà-shù cannot be modified by adverbs such as hěn ‘very’, fēicháng ‘extremely’, tǐng ‘quite, rather’, while it can be externally modified when used as an autonomous sign, cf. (4d-f). All the other complex signs in (2) also satisfy this criterion.

(4) a. *hěn dà-shù very be.big-tree ‘very big tree’ b. *fēicháng dà-shù extremely be.big-tree ‘extremely big tree’ c. *tǐng dà-shù very be.big-tree ‘quite big [a] tree’

d. hěn dà de shù very be.bigDE₁ tree e. fēicháng dà de shù extremely be.big DE₁ tree
f. tǐng dà de shù ‘quite big [a] tree’

1.2.2 N₁-N₂ expressions

The complex signs in (3) are N₁-N₂ compound wordforms (e.g. yǔyán-yánjiù language-study lit. ‘language study’ = ‘study on (a) language(s)) as becomes apparent when the external modifier criterion is applied to one of them. In (5a), yǔyán- as a constituent of yǔyán-yánjiù cannot be modified by an external element such as qíguài ‘be.weird’, while it can be externally modified when used as an autonomous sign as in (5b). The other complex signs in (3) also satisfy this criterion.

(5) a. Zhè shì yī zhòng qíguài de yǔyán-yánjiù. This be one kind be.weird DE₁ language-study This sentence necessarily means ‘It is a weird study on languages.’ and not ‘It is a study on [a] weird language(s).’ b. qíguài de yǔyán be.weirdDE₁ language ‘[a] weird language(s)’

⁶ In (Huang, 1984, 60), it is argued that most of the differences between a wordform and a phrase can be reduced to the Lexical Integrity Hypothesis (LIH): “No phrase-level rule may affect a proper subpart of a word.”

⁷ It can be observed that yǔyán- and -yánjiù are themselves compound stems.
1.3 The non-conjunction reduction criterion

The non-conjunction reduction criterion can be formulated as follows: \( S_1S_2 \) and \( S_3S_2 \) are two complex signs, conjoined in a construction such as “\( S_1S_2 \) Coordinate Conjunction \( S_3S_2 \)”. \( S_1-S_2 \) and \( S_3-S_2 \) are wordforms if \( S_2 \) cannot be factorized to create a construction like “\( S_1 \) Coordinate Conjunction \( S_3 \)”. This criterion has been proposed in (Huang, 1984) under the name of conjunction reduction criterion. It has also been mentioned in (Fan, 1958) and (Duanmu, 1998).

1.3.1 QV-N expressions

The second criterion confirms that \( \text{dà-shù} \) in (2a) is a QV-N compound wordform. In (6a), where \( \text{dà-shù} \) is conjoined with \( \text{xíao-shù} \) be.small-tree ‘small tree’, it is not possible to factorize \(-\text{shù}\), cf. (6b). But when \( \text{shù} \) is used as an autonomous sign in a coordinate construction, as in (7a), it can be factorized, cf. (7b):

(6) a. \( \text{Wǒ xǐhuān xíao-shù hé dà-shù.} \) me like be.small-tree and be.big-tree ‘I like small trees and big trees.’

b. \( \ast \text{Wǒ xǐhuān xíao hé dà-shù.} \) me like be.small and be.big-tree ‘I like small and big trees.’

(7) a. \( \text{Wǒ xǐhuān xíao de shù hé dà de shù.} \) me like be.small DE1 tree and be.big DE1 tree ‘I like small trees and big trees.’

b. \( \text{Wǒ xǐhuān xíao de hé dà de shù.} \) me like be.small DE1 and be.big DE1 tree ‘I like small and big trees.’

1.3.2 N₁-N₂ expressions

The second criterion also confirms that \( \text{yúyán-yánjìù} \) in (3a) is a N₁-N₂ compound wordform. In (8a), where \( \text{yúyán-yánjìù} \) is conjoined with \( \text{mínzú-yánjìù} \) nation-study ‘study on [a] nation(s)’, it is not possible to factorize \(-\text{yánjìù}\), cf. (8b). But when \( \text{yánjìù} \) is used as an autonomous sign in a coordinate construction, as in (9a), it can be factorized, cf. (9b):

(8) a. \( \text{mínzú-yánjìù hé yúyán-yánjìù} \) nation-study and language-study ‘[a] study on [a] nation(s) and [a] language(s)’

b. \( \ast \text{mínzú hé yúyán-yánjìù} \) nation and language-study ‘[a] study on [a] nation(s) and [a] language(s)’

---

8 Another version of this criterion is: \( S_1S_2 \) and \( S_1S_3 \) are two complex signs conjoined in construction such as “\( S_1S_2 \) Coordinate Conjunction \( S_1S_3 \)”. \( S_1-S_2 \) and \( S_1-S_3 \) are wordforms if \( S_1 \) cannot be factorized in such a way as we have a construction like “\( S_1 \) S₂ Coordinate Conjunction S₃”.
(9) a. mǐnzú de yànjiū hé yǔyán de yànjiū
    nation DE₁ study and language DE₁ study
    ‘[a] study on [a] nation(s) and [a] study on[a] language(s)’

   b. mǐnzú de hē yǔyán de yànjiū
    nation DE₁ and language DE₁ study
    ‘[a] study on [a] nation(s) and [a] language(s)’

It is easy to show that the other complex signs in (2)-(3) also satisfy this criterion. We will not do so for lack of space.

In the case of N₁N₂ complex signs, a third test could be added: that of “the anaphoric island”, which can be formulated as follows: “If nominal stems N₁- and -N₂ constitute a wordform N₁-N₂, it should not be possible to use an anaphoric pronoun to refer to the referent of N₁ in one sentence.” But this test does not seem to be so effective when applied to Chinese. For example, even if (10) is deemed awkward and unnatural, it does not really back up our hypothesis that yǔyán-yànjiū is a wordform since (11), where yǔyán de yànjiū is a phrase, is also deemed awkward by our informants. To sound natural (10) and (11) should be respectively corrected into (12) and (13):

(10) Zhè zhǒng yǔyán-yànjiū hěn tèbié, tā jiù shì hànyǔ.
    this kind language-study very be.special it exactly be Chinese
    ‘This language study is special, it is exactly Chinese.’

(11) Zhè zhǒng yǔyán de yànjiū hěn tèbié, tā jiù shì hànyǔ.
    this kind language DE₁ study very be.special it exactly be Chinese
    ‘This language study is special, it is exactly Chinese.’

(12) Zhè zhǒng yǔyán-yànjiū hěn tèbié, tā jiù shì hànyǔ-yànjiū.
    this kind language-study very be.special it exactly be Chinese-study
    ‘This language study is special, it is exactly a study on Chinese.’

(13) Zhè zhǒng yǔyán de yànjiū hěn tèbié, tā jiù shì hànyǔ de yànjiū.
    this kind language DE₁ study very be.special it exactly be Chinese DE₁ study
    ‘This language study is special, it is exactly a study on Chinese.’

2 Collocational compositional compounds [= CCC]

Among compositional compounds, some are totally free, i.e. they do not show any phraseologization and are created freely and productively by a speaker, such as xiǎo-huā be.small-flower ‘small flower’ and mùtou-fāngzi wood-house ‘wooden house’. Others are partially phraseologized and constitute morphological semi-phraseomes or collocational compositional compounds: the combinatorial properties of one of their components are affected by some phraseologization.
2.1 N₁-N₂ collocational compositional compounds

Since I already examined the case of a QV-N collocational compound, cf. 大雨 be.big-rain ‘heavy rain’ (see Introduction), I will turn to two examples of N₁-N₂ collocational compounds: 房门 house-door ‘house door’ and 午餐 noon-meal ‘lunch’.

房门 house-door ‘house door’ is a CCC for the following three reasons:

- It is a compound wordform; it satisfies the two criteria for wordhood mentioned above.
- Its meaning is compositional (the sum of its constituents’ signified is equal to its signified: ‘房- [= house] ⊕ ‘-门 [= door]’ = ‘房门 [= house door]’).
- The selection of the stem 房- is restricted. It would be impossible to say 房子门 house-door ‘house door’, cf. (14a-b), even though 房- and 房子- have the same meaning (房子- is a derived stem, obtained by suffixing the semantically empty suffix -子 to the stem 房-)

(14) a. *这裡的 房子门 很 大。 b. 这裡的 房门 很 大。
here DE₁ house-door very be.big ‘House doors are very big here.’

The other example, 午餐 noon-meal ‘lunch’ is a CCC because it satisfies the same three criteria as 房门 regarding its wordhood, its semantic compositionality and collocational restriction.

- It satisfies the two criteria for wordhood mentioned above.
- The sum of its constituents’ signified is equal to its signified: ‘午- [= noon]’ ⊕ ‘-餐 [= meal]’ = ‘午餐 [= lunch]’.
- The selection of the stem 午- is dictated by the language. It is not possible to say for example 中餐 with the meaning ‘lunch’.

Here are a few more CCCs I identified: 早饭 morning-meal ‘breakfast’, 午饭 noon-meal ‘lunch’, 晚饭 evening-meal ‘supper’, 餐厅 meal-room ‘dining room’, 宴饭 be.full-meal ‘a meal that satisfies one’s appetite’, 大觉 be.big.sleep ‘long sleep’, 午睡 noon-sleep ‘noon sleep’, 群舞 group-dance ‘group dance’, 热水澡 be.hot.water-wash ‘hot shower or hot bath’.

---

9 The wordform 中餐 exists in Chinese. But its meaning is ‘Chinese food’. 中- ‘China’ should not be confused with 中- ‘noon’, which can be found in 中饭 noon-meal ‘lunch’.

10 The selection of the stem 热水 is restricted. It is not possible to say * 热zào be.hot-wash, *热 de zào be.hot DE₁ wash.
2.2 Lexicographical treatment of CCCs

2.2.1 Lexical unit or not?

The existence of such compounds raises a lexicographic question: how should one present CCCs in an Explanatory and Combinatorial Dictionary of Chinese? They are neither free compounds, nor are they non-compositional compounds. Should they be regarded as lexical units and treated in separate dictionary entries? Using the analogy of collocations, I think that they should not. A collocation does not merit an entire lexicographic article because all the information necessary for its proper use is given where the collocation is introduced, in the combinatorial zone of the key lexeme. For example, it would not be useful to treat the collocation to rain cats and dogs as a separate entry if all that can be said about this expression is that cats and dogs means ‘intensely’ when used with rain. This information is given in the lexical combinatorial zone of the lexeme RAIN by the lexical function \[= LF\] \text{Magn: Magn}(\text{rain}) = \text{cats and dogs, buckets, heavily, etc.} To differentiate CCC lexemes from real lexical units I will therefore write them with a different font (Arial). For example, zǎo-cǎn morning-meal ‘breakfast’, which is a CCC will be noted in Arial, and ZÀO-DIÀN morning-dot ‘light breakfast’ or YÈ-XIÀO night-night ‘midnight snack’, which are NCC and thus lexical units, in Times New Roman.

2.2.2 Collocations versus Collocational Compositional Compounds

Another important question concerns the differentiation of CCCs from real collocations in the dictionary. Two solutions can be considered.

The first one consists in presenting a CCC as a fused element of the value of a lexical function. For example, dà-yǔ would be written as follows: \text{Magn}(yǔ) = // dà-yǔ be.big-rain ‘heavy rain’. The symbol // before dà-yǔ indicates that it is a “fused” element. A fused element cumulatively expresses the meaning of an LF (here ‘\text{Magn}’) and its keyword ‘yǔ’, so that the latter is not repeated in the sentence. The problem with this notation is that it hides the fact that dà-yǔ has a compositional meaning. It makes dà-yǔ appear like a NCC such as downpour, which is not the case.

The second solution would be to introduce a CCC the same way as a real collocation. For example, in a traditional ECD, collocations like fierce battle or jǐliè de zhàn-dòu be.fierce DE₁ battle ‘fierce battle’ would be presented as follows: \text{Magn(battle) = fierce} ~. \text{Magn(zhàn-dòu) = jǐliè de} ~. A CCC like bǎo-fàn be.full-meal ‘meal that satisfies one’s appetite’ could be presented the same way as above: [\text{Bon + Magn}(fàn) = bǎo ~. But this notation tends to hide the fact that bǎo-fàn is a compound wordform. Without any clue of its morphological status, learners of Chinese could be induced to regard bǎo-fàn as a phrase. As a consequence, they are likely to create ungrammatical sentences such as (15):

(15) *Wǒmén chī hén bǎo de bàn.  
us eat very be.full DE₁ meal  
‘We eat a very satisfying meal.’

---

\text{Mel’čuk et al., 1995), and (Mel’čuk et al., 1998).}
Therefore I suggest that a CCC should be written as follows: \([\text{Magn} + \text{Bon}] (\text{fàn}) = \text{bǎo[fàn]}\). In this way, learners will know that \(\text{bǎo[fàn]}\) is a compound lexeme. The brackets surrounding \(\text{fàn } \text{‘meal’}\) also shows that the element of the value of the LF \([\text{Magn} + \text{Bon}] (\text{fàn})\) is \(\text{bǎo}\) only.

Sometimes, a collocation and a CCC made of the same constituents can coexist in Chinese. This is the case with \(\text{dà-yǔ} \text{‘heavy rain’}\) and \(\text{dà de yǔ} \text{‘big DE1 rain’}\). They should be presented as follows: \(\text{Magn}(\text{yǔ}) = \text{dà de } \sim, \text{dà[yǔ]}\).

**Conclusion**

Chinese does have semantically compositional compound wordforms and some CCs are morphological semi-phrasemes. To our knowledge, no studies so far have ever discussed this collocational phenomenon, even though it seems to be widespread: indeed CCCs are abundant in Chinese. The fact that collocational phenomena occur at both syntagmatic and morphological levels might partly explain why it has never been easy for Chinese lexicographers to define clearly what constitutes a collocation in this language. CCCs merit in-depth investigation for several reasons:

- A thorough study of these morphological semi-phrasemes can help us gain a better understanding of composition in Chinese and in other languages.
- Understanding these compounds would improve lexicographic treatment of collocations in Chinese dictionaries, where usually no clue is given to learners as to whether an expression is a lexical unit or a CCC, a real collocation or simply an idiom.
- The study of CCCs would provide us with a broader view of phraseology, which concerns wordforms as well as phrases. The existence of CCCs suggests a typology of compound words (i.e. morphological full phrasemes, quasi-phrasemes, semi-phrasemes) that parallel to that of phrases, see (Mel’čuk, 1998, 2003).

**Acknowledgements**

I am particularly grateful to Prof. Igor Mel’čuk for his invaluable comments and suggestions. I would also like to thank Noah Constant who thoroughly proofread an early draft of this paper, and my third year students from Hebei University in China who served as great linguistic informants. Finally I wish to thank Josh Holden for rigorously proofreading the final draft.

**Bibliography**


Étienne Van Tien Nguyen


