Morpho-Syntax of Non-VO Separable Compound Verbs in Cantonese

Sheila S.L. Chan, Lawrence Y.L. Cheung
The Chinese University of Hong Kong

Abstract

Verb-object (VO) separable compound verbs (SCVs), for example, *lí-le-hūn* 'divorced', have long been studied. A small group of non-VO compounds in Cantonese are also separable, but have not yet been addressed. In this study, a preliminary judgment test was used for the first time, to look into the separation of non-VO compounds. We found that the separation of non-VO compounds, though limited, is different from that of VO compounds in terms of their ways of separation. There seems to be an effect of the ways of separation and the morphological structures of the verbs on the separability. We also showed that the underlying identity of non-VO SCVs is lexical, as most of them do not have a phrasal form. This group of separable verbs, which was neglected before, could have an impact on related morpho-syntactic theories.

Keywords

morpho-syntax, separable verbs, Cantonese

1 Background

The separable compound verb (SCV) is a long-standing puzzle in Chinese linguistics. Compound verbs in Chinese consist of two morphemes, for example, *líhūn* separate + marriage (= 'divorce') and *lǐfà* manage + hair (= 'have a haircut'). A subset of the compound verbs, primarily those with a verb-object (VO) morphological structure, has been claimed to display both lexical and phrasal properties. On the one hand, many SCVs are semantically non-compositional. In other words, the meaning of many SCVs cannot be derived entirely from the meanings of its constituent morphemes (Li and Thompson 1989). For example, the meaning of *shāngfēng* 'catch a cold' is not directly a combination of the meaning of its constituents *shāng* 'hurt' and *fēng* 'wind'. The non-compositionality of SCVs implies that it is better to treat them as an integral whole lexically. On the other hand, most of these VO compounds are separable by various elements and their meanings are retained even after separation, suggesting that the V and O are related syntactically like a verb phrase. Although previous studies focus mostly on SCVs in Mandarin, a similar phenomenon is also found in Cantonese.

Despite the previous research on VO SCVs in Mandarin, there remain some major research gaps. Some scholars have noted that non-VOs are also separable to some extent (Chao 1968; Fu 2016; Li and Thompson 1989; Zhou 2006). However, systematic investigation of non-VO SCVs is lacking. Our preliminary observations are that even compounds of the same morphological type vary a lot in their degree of separability. For example, while most VO compounds are separable, others are not. This phenomenon is even more prominent when it comes to non-VO compounds. Second, there is some controversy in the research on whether SCVs are basically lexical or phrasal. Previous analyses of SCVs (Chao 1968; Huang 1984; Packard 2000; Xie 2001; Zhou 2006) are built primarily on data from VO SCVs, without little consideration of non-VO SCVs. We will see later that these analyses run into problems when they are applied to non-VO SCVs. Finally, this paper will investigate SCVs in Cantonese, as separable non-VOs seem to be not uncommon in Cantonese. The findings from Cantonese will most likely shed light on the general analysis of SCVs in Chinese.

To fill the above research gaps, this study addresses two research questions. First, how separable are non-VO compound verbs compared with VO compound verbs? Understanding the properties of non-VO SCVs shows how different they are from VO SCVs. Second, are SCVs basically lexical or phrasal? A better analysis of SCVs, including both VO and non-VO, could help inform the morphological theories of how syntax and morphology interact, as SCVs seem to straddle between the two. The rest of the paper is organized as follows. Section 2 reviews the general properties of SCVs in Chinese. Section 3 examines closely the separability of SCVs by various elements. Based on the findings, Section 4 presents our analysis of SCVs. Finally, the conclusion is drawn in Section 5.

2 Separable compounds

2.1 VO versus non-VO

Chinese compound verbs can be classified into five major types according to the morpho-syntactic relationship between the morphemes in the compound. Besides VO, there are four types of compounds: subject-predicate (SP), coordinate (CO), subordinative (SB), and verb-resultative (VR) (Chao 1968; Lu 1964). It has been widely noted that the two morphemes of most VO compounds are separable by aspect markers, measure phrases, and modifiers of the object constituent, as in (1)–(4). Some objects can even undergo movement to a position before the verb constituent (Li and Thompson 1989), as in (4) below:

(1) Tā   <u>lù</u>   -le   -<u>yīn</u>.
     he   record   -PERF   -sound
     'He has recorded the sound.'

(2) Wǒ   <u>bǔ</u>   -le   -liǎng-cì   -<u>zhuāng</u>.
     I   reapply   -PERF   -twice   -makeup
     'I have reapplied the makeup twice.'

(3) Nǐ   <u>bāng</u>   -tā   -de   -<u>máng</u>.
     you   help   he   -GEN   -busy.
     'You help him.'

(4) Wǒ   lián   <u>wǔ</u>   dōu   bù   huì   <u>tiào</u>.
     I   even   dance   all   not   know   jump
     'I don't even know how to dance.'

In contrast, non-VO types are far less separable.[1] The separation strategies mentioned in (1)–(4) often do not apply (see examples in Column [b] in Table 1 with Cantonese perfective aspect marker *zo2*). However, previous studies (Chao 1968; Li and Thompson 1989) note that they are separable to some degree (see examples in Column [d] in Table 1).

The present research will provide empirical evaluation of the differences in separability between VO and non-VO compounds.

---

1 Some VR compounds are separable by potential affixes *de* 'able' and *bu* 'not', but in that case, they are generally not regarded as a compound but a VR phrase (Chung 2006), such as *shuì-dé-zháo* 'fall asleep' and *shuì-bù-zháo* 'cannot fall asleep'. Some could even be separated by negation markers in Cantonese, like *sau6-m4-lok6* 'cannot accept'. A few SP compounds are also separable by intensifiers, negation markers, and adverbs (Li and Thompson 1989), like *xīn-hěn-jí* 'very impatient', *xīn-bù-jí* 'not impatient', and *tā xīn-chángcháng-jí* 'he is often impatient'. In the samples we collected, coordinate and subordinative compounds are also found separable sometimes.

Table 1 Examples of non-VO compound verbs and their separability by *zo2*

Morphological Type	(a) Verb	(b) Inseparable by *zo2*	(c) Verb	(d) Separable by *zo2*
1. SP	*zi6baau3* 自爆 'expose oneself'	**zi6-zo2-baau3*	*zi6sau2* 自首 'turn oneself in'	*zi6-zo2-sau2*
2. CO	*ziu1laam5* 招攬 'recruit'	**ziu1-zo2-laam5*	*ceot1faat3* 出發 'depart'	*ceot1-zo2-faat3*
3. SB	*am3lyun2* 暗戀 'have a crush'	**am3-zo2-lyun2*	*tung4geoi1* 同居 'live together'	*tung4-zo2-geoi1*
4. VR	*zing2ding6* 整定 'destine'	**zing2-zo2-ding6*	*fong3daai6* 放大 'magnify'	*fong3-zo2-daai6*

2.2 Basic properties of VO separable compounds

SCVs possess both lexical and phrasal properties. Although SCVs consist of two morphemes, these compounds are semantically non-compositional (Li and Thompson 1989). As mentioned before, the idiomatic meaning of *shāngfēng* 'catch a cold' could not be understood through the constituents *shāng* 'hurt' and *fēng* 'wind' directly. They also usually have at least one bound morpheme, like *gé* in *gémìng* 'revolt' and *lǐ* in *lǐfà* 'have haircut'. In this sense, the two constituents function together as a lexical morpheme.

SCVs at the same time can appear as a phrase-like structure and are seemingly subject to syntactic operations. Most VO compounds and a small group of non-VO compounds in Cantonese are separable by various elements with their idiomatic meanings still retained even after the separation. They would typically admit verbal suffixes like aspect markers, as in (5). Sometimes, modifiers of objects (6)–(7) and measuring phrases (8)–(9) could be added to the compound. Some highly separable compounds even allow the movement of the object constituent morpheme, such as the object preposing in the *lián…dōu* construction in (4).

(5) Tā   bì-le-yè.
     he   finish-PERF-study
     'He has graduated.'

(6) fèi   -le   -nín   -xǔduō   -shén
     separate   -PERF   your   many   energy
     'spent a lot of your energy'

(7) xuě   -zhège   -chǐ
     revenge   this   shame
     'revenge this wrong'

(8) bāng   -guò   -jǐ   cì   -máng
     help   -EXP   several   time   busy
     'give some help several times'

(9) lí   -guò   -sāncì   -hūn
     depart   -EXP   three.time   marriage
     'got divorced three times'

Most separable verbs are intransitive and cannot take a direct object. This may follow from the phrase structure condition (PSC) (Huang 1984), which states that a verb can only be followed by at most one constituent and disallows a VO compound to take an object. There are, however, some exceptions like *dānxīn* 'worry' and *zhùyì* 'pay attention to'. For example, if *dānxīn* [burden + heart] 'worry' has to take an object, the PSC should have ruled out *dānxīn tā* 'worry (about) him/her', because *dān* 'to burden' has admitted *xīn* 'heart' and should not be able to accommodate *tā* 'him/her'. To explain (10a), it is possible that *dānxīn* 'worry' as a

whole is reanalyzed as a complex verb and can admit one constituent *tā* 'him/her' without violating PSC. In (10b), in contrast, the compound is separated by *le* and cannot be subject to reanalysis. As the final morpheme *xīn* 'heart' becomes the object while the object *tā* is prohibited in the sentence as below:

(10) a. Wǒ   dānxīn   tā.
       I   worry   him/her
       'I worry about him/her.'

     b. *Wǒ   dān-le-xīn   tā.
         I   worry-PERF   him/her
         (Intended) 'I worried about him/her.'

2.3 Degree on separability

Although VO compounds are highly separable, the degree of separability of individual compounds varies. Their constituents have to "stay in the near context"; for example, they have to be in the same clause (Chao 1968). Chao (1968) suggests that the separability of SCVs can be tested with reference to whether they are as follows:

(a)   Admit suffixes and complements to the verb;
(b)   Admit modifiers to the object;
(c)   Allow inversion of the verb and object;
(d)   Allow separation of the verb and object in questions and answers.

However, there seems to be no general rule concerning the separation of VO compounds. Which compound can undergo what kind of separation has to be learned individually (Li and Thompson 1989).

      For example, a small group of VO compounds is completely inseparable (Li and Thompson 1989). Compounds like *yànshì* 'loathe the world' and *fēnmiǎn* 'give birth to a child' do not admit the separation of the morphemes by any of the strategies in (a)–(d). Some compounds like *shāngfēng* 'catch a cold' and *xǐzǎo* 'take a shower' are somewhat separable and can allow (a) and (b). Some other compounds, especially VO compounds, are highly separable, as they accept all four kinds of separation, like *jiéhūn* 'get married' and *bìyè* 'graduate'. According to Chao (1968), the object of these compounds can even be moved long-distance to the higher clause via topicalization, as in the Cantonese example in (11):
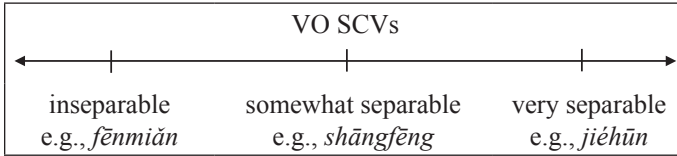
(11) Fan1   ne1,   ngo5   nam2   keoi5   dou1   git3   -maai4   laa3.
     Marriage   SFP   I   think   him   also   marry   -PRT   SFP
     'I think he has married.'

The different degree of separability is summarized in Table 2. We can see that the separability of SCVs is not homogeneous even among the VO type. Rather it is better to be conceived as a continuum, as illustrated in (12).

Table 2 Degree of separability of VO SCVs

Group	Example	(a)	(b)	(c)	(d)
1. Inseparable	*fēnmiǎn*	✕	✕	✕	✕
2. Somewhat Separable	*shāngfēng*	✓	✓	✕	✕
3. Highly Separable	*jiéhūn*	✓	✓	✓	✓

(12) Continuum of VO SCVs

```
                          VO SCVs
    ◄──────────┼──────────────┼──────────────┼────────►

     inseparable     somewhat separable     very separable
    e.g., fēnmiǎn     e.g., shāngfēng      e.g., jiéhūn
```

Chao (1968) highlights the existence of a small class of non-VO compounds which are very separable but are not VO structurally. Sometimes, pseudo-VO structures can even be a single morpheme like (13) and (14) as below:

(13) *yōumò* 'humorous' → *yōu-le-tā-yī-mò* 'make a joke with him'
(14) *kāngkǎi* 'generous' → *kāng-tārénzhī-kǎi* 'gen-with other people's-erous'

To explain this class, he tentatively suggests that the two morphemes of such special cases are "coerced" into a VO relationship and become a pseudo-VO compound (Chao 1968).

2.4 Non-VO SCVs in Cantonese

The discussion above is largely based on SCVs in Mandarin. Cantonese SCVs share similar properties. Examples of Cantonese SCVs are given below. Besides the five morphological structures, we have also included disyllabic verbs whose syllables do not display an obvious internal structure, for example, (19). We label them as "unclassified (structure)" in this study as follows:

● Coordinate (CO):
  jing2jan3 [take-a-shot+copy] 'photocopy',
  paak3to1 [accompany + hold-hand] 'date',
  bei2coi3 [compare+compete] 'compete'
(15) Bei2-jyun4-coi3   heoi3   sik6faan6.
     compete-finish   go   eat.rice
     'Go grab a meal after the competition.'

● Subordinative (SB):
  saan3bou6 [scattered+step] 'have a walk',
  ci4dou3 [late+arrive] 'be late',
  soeng1tai2 [mutual+see] 'blind-dating'

(16) Keoi5 gam1nin4   soeng1-zo2-saam1ci3-tai2.
   he   this.year   blind-date-PERF-three.times
   'He went to blind-dating three times this year.'

- Subject-Predicate (SP):
  hau2mat6 [mouth+tight] 'be tight-lipped',
  sam1taam5 [heart+thin] 'be disheartened',
  dei6zan3 [earth+shake] 'earthquake'

(17) Gam1nin4 Jat6bun2   dei6-gwo3-zan3.
   this.year   Japan   earthquake-EXP
   'Japan had an earthquake this year.'

- Verb-Resultative (VR):
  co5dai1 [sit+low] 'sit down',
  haak3zat6 [frightened+obstruct] 'be intimidated',
  daa2dai1 [fight+low] 'defeat'

(18) Keoi5 daa2-zo2-dai1 saam1baak3 go3 caam1gaa1ze2.
   he defeat-PERF   three-hundred CLASS participant
   'He defeated three hundred participants.'

- Unclassified (UN):
  zap1saang1 [pick+life] 'improvise',
  fei4lou2 [FAIL] 'fail',
  bong1can3 [help+match] 'frequent'

(19) Bong1-zo2-can3 gam3 noi6   ping4 di1 laa1.
   frequent-PERF so long-time cheap little SFP
   'Give me a discount as I have been your customer for a long time.'

2.5 Morpho-syntactic issues on SCVs

The discussion so far shows that many VO compounds and a small subset of non-VO compounds (especially in Cantonese) possess properties of a word and a phrase-like structure. Although the compounds are like lexical items, they could be manipulated by syntactic rules. They demonstrate indeterminacy or interaction between morphology and syntax (Packard 2000). Are they basically lexical or phrasal? There are three main proposals for the paradox of VO SCVs.

2.5.1 Ionization

All VO and VR SCVs are basically listed as words in the lexicon with reanalysis re-labelling them as phrases when appropriate (Chao 1968). Chao was the first one to refer to the phenomenon of non-VO SCVs as "ionization": "a compound can be expanded in a limited number of ways, as long as the constituents stay in the near context". This is potentially applicable to not only VO compounds, but also non-VO compounds. Even if the morphological relationship between the two morphemes is not VO, it could be coerced into a VO one. As the stress of a VO construction is always on the object, any iambic verbal expressions could

possibly be treated as VO and be "ionized" (Chao 1968). As seen in (13) and (14), constructions as a single morpheme could be "ionized" when they have the stress pattern the same as VO compounds.

2.5.2 Lexicalization

All VO and VR SCVs are basically listed as phrases in the lexicon (Huang 1984). Lexicalization will only be applied to the phrase in order to fulfill grammatical requirements. One such requirement is the PSC. For example, when the VO/VR has to admit an object (e.g., *dānxīn* + object), it first has to be lexicalized as a compound verb to observe the PSC. If no such need arises, the VO/VR will remain as a phrase. Huang (1984) prefers lexicalization to ionization because cross-linguistically, ionization as a rule of grammar appears to be extremely exotic but lexicalization is more common.

2.5.3 Dual listing

All SCVs can be listed in the lexicon as both words and phrases. Packard (2000) proposes a solution for the underlying lexical identity of VO forms. He claims that VO compounds can have a dual status in the lexicon as both words and phrases: When a VO form is lexicalized as a word for the first time, it is listed as a lexical item, and it would retain its phrasal form if both morphemes are free, for example, *kànshū* 'read'. If it is lexicalized to the point that it cannot be a phrase (one of the morphemes are not free anymore), it can still be reanalyzed as a phrase in limited ways (e.g., *shuìjiào* 'sleep'), regardless of its structure as virtually any disyllabic verbs can be separable. Meanwhile, Her (2010) also proposes that there are three types of VO sequences: Type I can only be used as a word, Type II as a phrase only, and Type III can be used as both a word and a phrase. Type III VO sequences which can be used as a word and a phrase are dual-listed in the lexicon and all the other types of VO sequences and non-VO compounds are only listed as words (or idioms for Type II VO sequences). The syntactic and semantic requirements of the item will also be listed together with the item in its entry. In this way, the grammar of VO sequences will not be complicated by those that can be used as both words and phrases. But this cannot account for the existence of separable verbs. The fact that non-VO compounds or Type I/II VO sequences are only listed as words or idioms cannot explain under what circumstances VOs can be separable.

It is obvious that the use of the phrasal form of SCVs is restricted. The phrasal form will only occur when the SCV is used with certain specific elements; on other occasions, the SCV will be used as a word and is inseparable. The constituents of the SCVs could not be used in the same way with any other morphemes or characters like free words do. We could see from the above theories that the underlying identity of SCVs is still controversial. As pointed out by Huang (1984), it is difficult to tell the underlying identity of SCVs based on VO compounds only because most VO compounds are separable. The three analyses are largely compatible with the data from VO SCVs, making it difficult to tease apart the analyses. In this study, we will look at non-VO compounds, which will shed light on the analysis, as the phrasal

form of non-VO compounds is not as readily available as that of VO compounds. It is necessary for us to understand, first, the properties of non-VO SCVs and their differences with VO SCVs, and second, the identity of non-VO SCVs and its relationship with existing morpho-syntactic theories.

3 Separability of non-VO compounds

To address the research questions, two studies have been designed.

3.1 Study 1

The goal of Study 1 is to examine the extent to which non-VO compounds are separable. We tested a range of elements that can potentially be inserted into each type of compounds. Disyllabic verbs in the two Cantonese dictionaries, *Xianggang Yueyu cidian* (Cheng 1997) and *Gang shi Guangzhouhua cidian* (Zhang and Ni 1999), were collected for constructing examples for the judgment tests. This study focuses on the separation of two-character words (not necessarily two morphemes): 445 non-VO compounds and 878 VO compounds were collected.

The collected examples of non-VO compounds were then analyzed morpho-syntactically. The morphological categories of the compounds are identified, for example, *ceot1faat3* 'depart' is classified as a coordinate (CO) compound. Since neither of the Cantonese dictionaries includes the usage or morphological category of the verbs, their category was determined by the authors who are native Cantonese speakers. As the nature of morphemes is one of the factors to be examined in this paper, the verb/noun-like nature of the morpheme is also identified with reference to *Xiandai Hanyu zidian* and *Xinhua zidian*. Each morpheme is looked up separately and the parts of speech (adjective, noun, or verb) listed in the entries of the morpheme in both dictionaries are marked (Figure 1). If the morpheme is not included in the dictionaries, its part of speech was judged by the authors.

To find out the separability of the compound verbs, the following methods are used.

3.1.1 Separability based on Google web search

A helpful way to find out whether a compound verb S_1S_2 (e.g., *daai4tou3* 大肚) can be separated by an element X (e.g., *zo2* 咗) is to check whether S_1XS_2 (e.g., *daai4-zo2-tou3* 大咗肚) actually occurs on the web. This can be done by searching strings

Compound	Category	PoS			Initial morpheme			Final morpheme		
		PoS-adj	PoS-n	PoS-v	I-adj	I-n	I-v	F-adj	F-n	F-v
八卦	SB	adj	n	v	adj	n	v	-	n	-
出發	CO	-	-	v	-	-	v	adj	-	v
用底	VO	-	-	v	-	-	v	-	n	-

Figure 1 Examples of compound and morpheme labels

like "大咗肚" and "小過息" on Google. Six separating elements were considered, namely, aspect markers (*zo2, gwo3, gan2*), modal particle (*dak1*), quantifier (*saai3*), and resultative particle (*jyun4*). Most returned examples are found in informal writing and conversations in online forums or blogs. An element is only considered acceptable to separate a verb when at least three distinct examples of the element separating verbs were found. The results can be found in Table A1(a) in Appendix.

3.1.2 Separability based on native speakers' judgment

One of the authors (Judge 1) and another native Cantonese speaker (Judge 2) were invited to judge whether it is acceptable to separate each of the collected compound verb by the six separating elements. If the judge considered it acceptable to separate a compound verb by a separating element (e.g., *daai4-zo2-tou3*), then he had to provide an example sentence in which the sequence occurs. It should be noted that as the judges were only given the verb and the separating element such as *daai4-zo2-tou3*, this task did not control for the linguistic contexts in which the sequence occurs. Despite the limitation, we still find the results helpful in providing us with an overall picture of acceptability. The results can be found in Table A1(b) and (c) in Appendix.

3.1.3 Results

Table 3 shows the percentage of SCVs in each category. A compound verb is counted as "separable" only when all three sources (Google and two judges) accept at least one way of separation, even if the elements inserted are different (Table 4). For example, *siu2sik1* 'recess' is considered separable because all three sources allow the compound verb to be separated by at least one element. In contrast, *bong1can3* 'to frequent a shop' has acceptable examples from two of the sources

Table 3 Overall separability of compounds

Category of non-VO	No. of verbs	No. of SCVs	% of SCVs
CO	111	18	16%
MP	11	2	18%
SB	90	18	20%
SP	26	2	8%
VR	175	82	47%
UN	32	5	16%
Total of non-VO	445	127	29%
Total of VO	878	543	62%
Total	1323	670	51%

Table 4 Examples of the determination process of a compound's separability

Compounds and examples		Separability
jing2jan3 **'to photocopy' (CO)**		Separable
Google	Keoi5 jing2-**zo2**-jan3 zau6 bei2 ngo5 jap6 zaap6. he photocopy-PERF then BEI me enter.gate 'He allows me to pass the gate after he photocopied (something).'	
Judge 1	Bun2 syu1 nei5 jing2-**zo2**-jan3 mei6? CLASS book you photocopy-PERF yet 'Have you photocopied the book yet?'	
Judge 2	Ni1 bun2 syu1 mui5jat1jip6 ngo5 dou1 jing2-**zo2**-jan3. DEM CLASS book every.one.page me also photocopy-PERF 'I photocopied every page of this book.'	
siu2sik1 **'to recess' (SB)**		Separable
Google	Keoi5dei6 ji5ging1 siu2-**gan2**-sik1. they already recess-PROG 'They have already been having recess.'	
Judge 1	Siu2-**jyun4**-sik1 soeng5 me1 tong4? recess-finish up what class 'What is the class after recess?'	
Judge 2	Daai6gaa1 siu2-**jyun4**-sik1 heoi3 zaap6deoi6. everyone recess-finish go gather.queue 'Everyone gathers together after recess.'	
bong1can3 **'to frequent a shop' (UN)**		Inseparable
Google	Ngo5 mou5 bong1-**gwo3**-can3 keoi5 maai5 je5. I NEG frequent-EXP him buy thing 'I didn't bought thing from him before.'	
Judge 1	Ni1 gaan1 pou3 ngo5 bong1-**gwo3**-can3. DEM CLASS store I frequent-EXP 'I bought things from this store before.'	
Judge 2	No acceptable example	

only. As a result, it is treated as inseparable. The percentage of SCVs is obtained by the total number of SCVs in the category ("No. of SCVs") divided by the total number of verb ("No. of verbs").

Only 29% of non-VO compounds are separable. Evidently, they are far less separable than VO compounds, which have a separability of 62%. It should be noticed that even VO compounds are not as separable as expected, based on our criterion. The results across the three sources are consistent in a way that the trend of the overall separability of non-VO compounds across the categories and elements is similar, that is the less separable categories show low separability mostly in all three sources and the more separable categories show higher separability, though Judge 2's overall acceptability of SCVs is lower (Table A1[c] in Appendix).

SP compounds are the least separable. Only 8% of them are separable and they are only compatible with the aspect marker *zo2* and the resultative particle *jyun4*. VR compounds are comparatively more separable among non-VO compounds with 47% of separability, and they are most compatible with the modal particle *dak1*.

3.2 Study 2

The goal of Study 2 is to look at the separability of the compounds in a more controlled way with stimuli selected before included. We wanted to find out how acceptable the stimuli are by more native Cantonese speakers. Six compound verbs from each of the five categories of non-VO compounds were selected from Study 1. Three VO compounds are included for comparison. The compounds and words are selected based on their separability in Study 1; they are generally more separable according to the results from Google search and the two judges. The items with VO compounds are for baseline reference. A total of 33 verbs were selected for the grammaticality judgment test. Nine ways of separation were tested in this judgment task, as shown below. In addition to the six separating elements used in Study 1, we also included frequency phrase, duration phrase, and the movement of the final morpheme/syllable.

- Three aspect markers: *zo2*, *gwo3*, *gan2*
- One modal particle: *dak1*
- One quantifying particle: *saai3*
- One resultative particle: *jyun4* (Matthews and Yip 1994)
- Frequency phrase, for example, *jat1ci3* 'one time'
- Duration phrase, for example, *jat1zan6* 'a moment'
- Movement of final morpheme/syllable, e.g., *paak3to1* 'date' → *to1 -dou1-paak3-zo2* 'even dated'

Since the number of stimuli is large (33 items with nine tests, making up 297 sentences for judgment), the tests are divided into 3 sets with 33 stimuli each.

- Set 1: ASP1-*zo2*, MOD-*dak1*, FRE
- Set 2: ASP2-*gwo3*, QUA-*saai3*, DUR
- Set 3: ASP3-*gan2*, RES-*jyun4*, TOP

Subjects were asked to give their judgments of one set of stimuli and avoid over-pondering. The scale of scores is from 1 to 5, where 1 is *very unacceptable*, 5 is *very acceptable*, and 3 is *neutral*; 45 subjects participated in the tests (15 for each set).

3.2.1 Results

Figure 2 shows that different types of non-VO compounds demonstrate varying separability: SB > CO > UN > VR > SP. It is a bit surprising that VR compounds are not that separable, even though their morphological structure is somewhat similar to VO compounds.
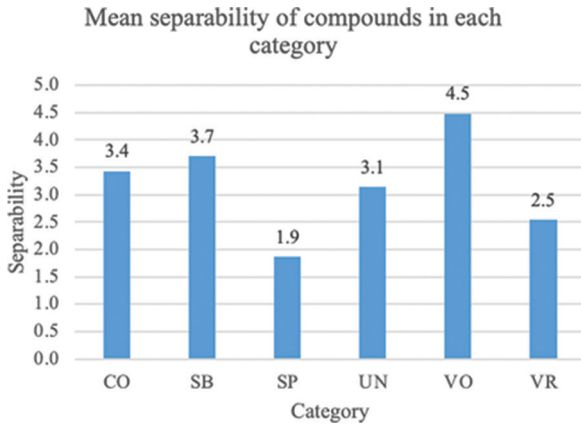
Mean separability of compounds in each category

Figure 2 Mean separability by morphological structure

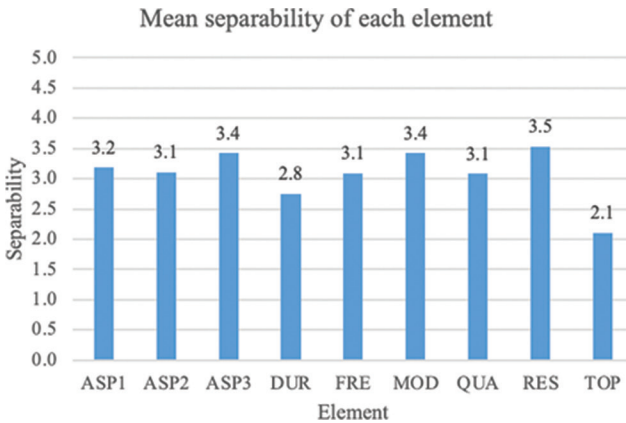Mean separability of each element

Figure 3 Mean separability by separating elements (non-VO compounds only)

Different elements of separation also display varying acceptability (Figure 3). The most acceptable one is the insertion of the resultative particle *jyun4*, as in *jing2-jyun4-jan3* 'photocopied' and *tung4-jyun4-geoi1* 'cohabitated'; while the least acceptable one is movement structure, like *lou2-keoi5-dou1-fei4-zo2* 'he failed' and *bou6-nei5-ji5-ging1-saan3-zo2* 'you have gone for a walk'.

There are also interactions between the categories of the compounds and the elements inserted, as shown in Table 5. VR compounds are more separable with the insertion of the modal particle *dak1*. While CO, SB, and UN compounds or words are very compatible with the three aspect markers, CO and SB compounds are more separable with experiential and progressive aspect markers *gwo3* and *gan2*, but UN compounds or words have a higher separability with the perfective aspect marker *zo2*.

Table 5 Mean of separability of categories and elements tested in Study 2

Category	ASP1 *zo2*	ASP2 *gwo3*	ASP3 *gan2*	DUR	FRE	MOD	QUA	RES	TOP	Mean
						Element				
CO	3.5	3.6	4.1	3.0	3.4	3.7	3.3	4.0	2.3	**3.4**
SB	3.7	4.0	4.0	3.6	3.9	4.1	3.4	4.1	2.6	**3.7**
SP	1.8	1.8	2.4	1.7	1.7	1.5	1.9	2.6	1.3	**1.9**
UN	3.7	3.1	3.1	2.6	3.4	3.3	3.7	3.5	1.9	**3.1**
VR	2.4	2.4	2.9	4.2	4.7	4.6	4.3	4.8	1.3	**2.5**
Mean	**3.2**	**3.1**	**3.4**	**2.8**	**3.1**	**3.4**	**3.1**	**3.5**	**2.1**	**3.1**

4 Discussion

4.1 Morpho-syntactical differences between VO and non-VO SCVs

Based on the findings from Sections 3.1 and 3.2, we highlight four differences between VO and non-VO compounds. First, non-VO compounds do not have the same degree of separability as VO compounds: the former are generally less separable. But one should notice that even VO compounds are less separable as expected. Second, even if the non-VO compounds are separable, each type is compatible with different sets of separating elements. Most can only accept aspect markers. Third, the element separating the two morphemes cannot be too large. As one can see, duration and frequency phrases, which consist of more than one syllable and one constituent, are generally not allowed for insertion. Finally, the movement of the final morpheme/syllable is totally banned in non-VO compounds. The fact that most non-VO compounds cannot be separated in ways like VO compounds suggests that non-VO compounds do not have a basic phrasal identity. This makes it difficult to assume that non-VO SCVs are basically phrasal and the lexical properties are derived via lexicalization of the phrasal structure (see Section 2.5.2).

One may argue that the separability of compound verbs is determined by the nature of the constituent morphemes. For example, if the constituent morphemes are free morphemes, then they are more likely to occur as a syntactic word and thus the structure can be a phrase (Packard 2000). Nevertheless, this does not seem to be the case when we look at non-VO compounds, for example, *jing2jan3* 'photocopy' and *ci4dou3* 'be late'. Both morphemes in the compounds are free and they are equally separable. In the example of *jing2seoi1* 'bring bad reputation' and *daa2lat1* 'get rid of", though both morphemes in the compounds are free, the compounds are not separable. We could see that whether the morphemes are free or bound is not the determining factor of a non-VO compound's separability.

4.2 Underlying identity of non-VO SCVs

As all the analyses of SCVs presented in Section 2.5 cover primarily VO compounds, we will first discuss to what extent the analyses can be extended to non-VO compounds. The assumption is that they can undergo similar, though limited, separation as VO SCVs do, and both the separation of VO and non-VO compounds can be viewed as a spectrum with a range of separability. We could look into the identity of non-VO compounds along three dimensions.

First, the results contrast with the lexicalization approach supported by Huang (1984), which states that VO or VR compounds are listed primarily as phrases in the mental lexicon. Many VO and VR compounds can undergo lexicalization to become lexical verbs. Although most VO compounds are separable, there are still some inseparable like *jim3sai3* 'loathe the world' and *fan1min5* 'give birth'. Previous studies and our data show that not all VO compounds are separable (38% of VO compounds in our study are inseparable). The existence of such inseparable

VO compounds is puzzling on the lexicalization analysis because it should predict that all VO compounds should be separable. This indicates that VO compounds do not necessarily have the phrasal forms. Non-VO compounds further complicate the problem because they are even less separable. If Huang's solution is to apply to non-VO compounds, one would fail to explain why the majority of non-VO compounds are not separable.

Second, our results partly echo the ionization analysis, which claims that SCVs are stored basically as lexical items and can be "ionized" and decomposes into a verb and an object. In the results, one could see that the compounds, as predicted by Chao (1968), generally do not favor the insertion of frequency phrases, duration phrases, and topicalization of final morphemes. Recall that Chao (1968) claims that some non-VO compounds can be coerced into pseudo-VO compounds. As only a minority of non-VO compounds can be separated, the coercion rule is very limited. If the relationship between morphemes in a non-VO compound is coerced into a VO one, it should pattern with the VO compound. From the results, it can be seen that the separation of non-VO compounds is very limited and different from that of VO compounds. The findings highlight that even if one accepts the notion of coercion, the coercion mechanism does not fully coerce a non-VO SCV into a separable verb.

Finally, the results only partly echo dual listing favored by Packard (2000) and Her (2010), which suggests that SCVs can have a dual status: they are stored as both a lexical item and a phrase in the lexicon. It is quite obvious that the underlying identity of non-VO compounds is lexical, as most of them do not have a phrasal form, for example, *jan5hei2* 'lead to'; even if they have one, the use of phrasal form is very restricted, for example, *sat6zaap6* 'to work as an intern' is only separable by the resultative particle. For compounds that do not have a phrasal form, according to Packard (2000), they are lexicalized from a phrasal form to the point that they cannot be a phrase. The compounds can then undergo a "reanalysis" (in Packard's [2000] word) to be used as phrases again. But it is unclear under what conditions can "reanalysis" be triggered. "Dual listing" here is different from Her's (2010) proposal, which states that VO sequences that can be used as both words and phrases are stored in the lexicon with both identities directly, because most of the non-VO compounds do not have a basic phrasal form. It is not possible for the SCVs to have two entries, one as a word and one as a phrase, in the lexicon right from the start. Generally, the proposal of dual listing has two major weaknesses. It cannot explain why some non-VO or even VO compound verbs are not separable. Furthermore, it does not state under what circumstances a compound verb can be separable.

According to the results obtained from the judgment tests, the non-VO compounds would not be phrases basically because most non-VO compounds cannot form phrases with other elements, and most importantly, the second morpheme/character in the compound cannot be moved/topicalized. The low

separability indicates that those non-VO compounds cannot be phrases as they do not have a phrasal form right from the beginning. The results from our study seem to be in favor with the concept of ionization suggested by Chao (1968).

4.3 Further studies on SCV and related morphological theories

The unique properties of SCVs could possibly shed light on the debate between lexicalism and distributed morphology (DM). Traditionally, lexicalists claim that syntax manipulates words created in the lexicon, a place where sounds and meanings combine. Lexical integrity hypothesis (LIH) states that rules of morphology can only manipulate units at the word level at most, while rules of syntax cannot manipulate units below the word level (Lieber 2018). In other words, syntactic operations can only occur at the word level but not below, but the separation of the morphemes in SCVs calls LIH into question as morphological constituents seem to be subject to syntactic manipulation. On the contrary, in DM, lexicon is replaced by lists of atomic roots and atomic bundles of grammatical features, phonological forms, and special meanings of particular roots (Marantz 1997). Syntax would operate with the roots and bundles of features, meaning that it works on not only words but also morphemes and some grammatical items. In that case, SCVs are not required to distinguish between a lexical entity and a phrasal entity stored in some places, which matches our finding that they do not have such an identity originally. The fact that SCVs can appear as both words and phrases might demonstrate how morphology and syntax interact with each other.

5 Conclusion

Our study used Google search and the grammaticality judgment tests to systematically look into the separation of non-VO compounds, which have not been done before in previous studies. We found that the separation of non-VO compounds is limited when compared with that of VO compounds in terms of the ways of separation allowed. The acceptability of different kinds of separation has been examined for each morphological type. We have also shown that the underlying identity of non-VO SCVs is lexical as most of them do not have a phrasal form. Through this study, one can see that previous studies and theories on SCVs neglect the existence of non-VO SCVs and do not accommodate the variability of separation. If VO and non-VO SCVs are to be accounted for in a similar fashion, the theories should also address the unique properties of non-VO SCVs. This is worth exploring in further studies as it might shed light on the syntactic or morphological mechanisms in the formation of SCVs.

References

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

Cheng, Ting Au. 1997. *Xianggang Yueyu cidian* [Hong Kong Cantonese dictionary]. Nanjing: Jiangsu Jiaoyu Chubanshe [Jiangsu education press].

Chung, Karen Steffen. 2006. *Mandarin compound verbs*. Taipei: Crane Publishing Co., Ltd.

Fu, Ning-lu. 2016. On the judgment of the separable words in contemporary Cantonese: A case of Cantonese phrase "Jung Mat Ji". *Journal of Guangdong Polytechnic Normal University* 6. 52–59.

Her, One-Soon. 2010. *Interaction and variation in the Chinese VO construction*. Taipei: Crane Publishing Co., Ltd.

Huang, C.-T. James. 1984. Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association* 19(2). 53–78.

Li, Charles N. & Sandra A. Thompson. 1989. *Mandarin Chinese: A functional reference grammar.* Berkeley, CA: University of California Press.

Lieber, Rochelle. 2018. Theoretical issues in word formation. In Jenny Audring & Francesca Masini (eds.), *The Oxford handbook of morphological theory*, 34–55. Oxford: Oxford University Press.

Lu, Zhiwei. 1964. *Hanyu de goucifa* [Morphology of Chinese]. Beijing: Kexue Chubanshe [Science press].

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *Penn Working Papers in Linguistics* 4. 201–225.

Matthews, Stephen & Virginia Yip. 1994. *Cantonese: A comprehensive grammar.* New York: Routledge.

Packard, Jerome L. 2000. *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Xie, Yaoji. 2001. The separation and reunion of words and phrases in Chinese. *Journal of Yantai University (Philosophy and Social Science)* 14(2). 234–240.

Zhang, Liyan & Liehuai Ni. 1999. *Gang shi Guangzhouhua cidian* [Dictionary of Cantonese in Hong Kong style]. Hong Kong: Wan Li Book Co., Ltd.

Zhou, Shangzhi. 2006. *Hanyu liheci yanjiu* [Research on separable words in Chinese]. Shanghai: Shanghai Waiyu Jiaoyu Chubanshe [Shanghai foreign language education press].

Appendix

Table A1 Separability of compounds from Google and two judges in Study 1 (Token in bracket)

(a) Google	Element					
Category	ASP zo2	ASP-*gwo3*	ASP-*gan2*	MOD-*dak1*	QUA-*saai3*	RES-*jyun4*
CO	17% (19)	14% (16)	14% (15)	9% (10)	16% (18)	15% (17)
MP	18% (2)	1 (9%)	18% (2)	18% (2)	27% (3)	9% (1)
SB	18% (16)	18% (16)	21% (19)	8% (7)	11% (10)	22% (20)
SP	12% (3)	4% (1)	4% (1)	0% (0)	4% (1)	12% (3)
VR	13% (22)	7% (13)	17% (29)	46% (80)	20% (35)	13% (23)
UN	25% (8)	13% (4)	16% (5)	6% (2)	25% (8)	13% (4)
Total of non-VO	16% (69)	11% (51)	16% (71)	23% (101)	17% (75)	15% (68)
VO	75% (659)	64% (562)	73% (638)	44% (387)	74% (648)	69% (609)
(b) Judge 1	Element					
Category	ASP-zo2	ASP-*gwo3*	ASP-*gan2*	MOD-*dak1*	QUA-*saai3*	RES-*jyun4*
CO	25% (28)	15% (17)	17% (19)	12% (13)	17% (19)	16% (18)
MP	27% (3)	9% (1)	18% (2)	18% (2)	18% (2)	9% (1)
SB	27% (24)	22% (20)	28% (25)	8% (7)	12% (11)	24% (22)
SP	12% (3)	4% (1)	4% (1)	4% (1)	0% (0)	12% (3)
VR	28% (49)	12% (21)	18% (32)	56% (98)	19% (34)	14% (24)
UN	28% (9)	16% (5)	22% (7)	6% (2)	22% (7)	13% (4)
Total of non-VO	26% (116)	15% (65)	19% (86)	28% (123)	16% (73)	16% (72)
VO	83% (725)	70% (614)	77% (680)	48% (419)	75% (662)	74% (650)
(c) Judge 2	Element					
Category	ASP-zo2	ASP-*gwo3*	ASP-*gan2*	MOD-*dak1*	QUA-*saai3*	RES-*jyun4*
CO	12% (13)	5% (6)	10% (11)	4% (4)	4% (4)	13% (14)
MP	27% (3)	9% (1)	9% (1)	0% (0)	9% (1)	18% (2)
SB	14% (13)	13% (12)	19% (17)	4% (4)	6% (6)	12% (11)

(*Continued*)

Table A1 Continued

(c) Judge 2	Element					
Category	ASP-*zo2*	ASP-*gwo3*	ASP-*gan2*	MOD-*dak1*	QUA-*saai3*	RES-*jyun4*
SP	4% (1)	0% (0)	0% (0)	0% (0)	0% (0)	4% (1)
VR	14% (24)	5% (9)	13% (23)	35% (61)	13% (22)	9% (15)
UN	3% (1)	0% (0)	9% (3)	3% (1)	9% (3)	3% (1)
Total of non-VO	12% (55)	6% (28)	12% (55)	16% (70)	8% (35)	10% (44)
VO	36% (317)	8% (69)	40% (355)	2% (20)	33% (292)	22% (189)
All	Element					
Category	ASP-*zo2*	ASP-*gwo3*	ASP-*gan2*	MOD-*dak1*	QUA-*saai3*	RES-*jyun4*
CO	11% (12)	4% (4)	5% (6)	1% (1)	2% (2)	8% (9)
MP	18% (2)	9% (1)	9% (1)	0% (0)	9% (1)	0% (0)
SB	9% (8)	10% (9)	13% (12)	1% (1)	3% (3)	9% (8)
SP	4% (1)	0% (0)	0% (0)	0% (0)	0% (0)	4% (1)
VR	6% (11)	3% (5)	9% (16)	24% (42)	9% (16)	7% (12)
UN	3% (1)	0% (0)	3% (1)	3% (1)	9% (3)	0% (0)
Total of non-VO	8% (35)	4% (19)	8% (36)	10% (45)	6% (25)	7% (30)
VO	34% (299)	12% (67)	49% (324)	2% (14)	43% (276)	45% (175)

Sheila S.L. Chan
Mailing address: Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, Hong Kong
Email: sheilachan@cuhk.edu.hk

Lawrence Y.L. Cheung
Mailing address: Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, Hong Kong
Email: yllcheung@cuhk.edu.hk

粵語非動賓離合詞的構詞句法

陳樞嵐、張欽良

香港中文大學

提要

關於動賓離合詞（例如"離了婚"）的研究一直有很多，但在粵語中有少部分非動賓複合詞可以被分離使用，卻又並未有人深入研究。本研究初次使用判斷測試去探討非動賓複合詞的分離用法。我們發現這種用法有很多限制，同時亦可看出其與動賓離合詞的差異。分離的形式及複合詞本身的結構對詞語可分離程度有一定影響。我們亦展示出非動賓離合詞的本質身分是詞，它們大部分都沒有短語的形式。這類離合詞長期被忽視但相信對相關構詞句法理論會有影響。

關鍵詞

構詞句法，離合詞，粵語