

Jiajin Xu and Xiaochen Li

Structural and semantic non-correspondences between Chinese splittable compounds and their English translations: A Chinese-English parallel corpus-based study

Abstract: Splittable Compounds (SCs henceforth) are a rather productive type of disyllabic verbal construction in Chinese, whose two morphological elements can be used together or separated by other interposing elements while the semantic integrity remains. The present study examines the English translations of SCs in a Chinese-English parallel corpus of five million Chinese characters and English words. The study takes a form-meaning combined approach to the structural and semantic differences or non-correspondences between SCs and their English translations. The non-correspondences of lexico-grammar and argument structure of SCs and their translations were thoroughly annotated and analysed. The predominant pattern revealed by our corpus investigation is that the English translations of SCs are characterised by semantic explicitation as well as structural explicitation. This is especially true in the English translations of non-split SCs. The split and non-split SCs are translated in different ways morpho-syntactically and semantically. The translations of non-split SCs more often involve rearrangements in argument structure, such as argument addition and semantic alternative, while the translations of split SCs are more likely to involve changes in grammatical properties. The reconfiguration of morpho-syntactic elements and semantic roles may well be explained from the typological differences of the two languages, the verb semantics, and the syntactic contexts of the SCs.

Keywords: Splittable Compounds (SCs) in Chinese, parallel corpus, Chinese-English translation, morpho-syntax, argument structure, structural explicitation, semantic explicitation

Jiajin Xu: National Research Centre for Foreign Language Education, Beijing Foreign Studies University, Beijing, China. E-mail: xujiajin@bfsu.edu.cn

Xiaochen Li: School of Foreign Languages, Beihang University, Beijing, China
E-mail: rexhit@gmail.com

1 Introduction

Splittable Compounds (SCs henceforth), also known as 离合词 *liheci*, a term coined by Lu (1957: 22) which literally means ‘separate combined words’,¹ are a very productive type of disyllabic verbal construction in Chinese. The two morphological elements can be used together or separated by other interposing elements while the semantic integrity remains (e.g. 帮忙 *bangmang* ‘to help’ can be expanded into 帮了一个大忙 *bang-le yi-ge da mang* ‘offered a big helping hand’,² while the semantics of ‘to help’ is retained) (cf. Siewierska et al. 2010).

The objective of the present study is to investigate the structural and semantic non-correspondences between SCs and their English translations in a Chinese-English parallel corpus. A ‘non-correspondence’ is defined in this article as the addition, omission, or other type of structural or semantic adjustments in Chinese-English translation which yields differences between the source and the target text. The significance of studying the English translations of SCs on the basis of a parallel corpus is twofold. First of all, it offers a quantitative description of how Chinese SCs are translated into English. Secondly, from an applied point of view, the patterns of correspondences of SCs in translational texts will be valuable resources for human translators as well as for machine translation systems. Meanwhile, the results of the study will be a good aid in compiling Chinese-English bilingual dictionaries. In a nutshell, the cross-linguistic comparison between SCs and their English counterparts might open up more avenues for studies of Chinese morphology, lexicography, (machine) translation as well as the Teaching of Chinese as a Second Language.

This article is organised as follows. Section 2 briefly reviews the previous works of corpus-based studies on SCs. Section 3 describes the retrieval and the structural and semantic annotation of SC instances in the parallel corpus. Section 4 offers a descriptive analysis of the patterns of structural and semantic non-correspondences between 22 SCs and their English equivalents. Section 5 is an in-depth qualitative analysis of the non-correspondences of seven SCs and their

¹ When Lu proposed the new term 离合词 *liheci* in a footnote of his book *Hanyu de Goucifa* (Chinese Morphology), he did not actually refer to the specific type of verbal compounds discussed in the present study. He used the term to refer to cases like 羊肉 *yangrou* ‘mutton’ and 羊的肉 *yang de rou* ‘sheep’s meat’. In this context, 离合词 *liheci* is a composite term for 羊肉 *yangrou* and 羊...肉 *yang...rou*. Later, the term 离合词 *liheci* has been used in Chinese linguistics to exclusively refer to the verbal compounds in question.

² In this article, the first morphological element is notated as SC_H (splittable compound head), meaning the verb head of the SC, and the final element as SC_T (splittable compound tail), which is most likely to be a nominal morpheme.

English equivalents. The findings of Sections 4 and 5 are then discussed in Section 6, which compares the split and non-split SCs in the English translations. Section 7 concludes the research.

2 Corpus-based studies on SCs

Monolingual corpus based Chinese studies have a history of over thirty years in China (Feng 2006: 176). However, most of them have had emphases on applied linguistics, such as the compilation of Chinese wordlists and the application of corpora in machine translation systems. Corpus-based language descriptions have not attracted enough attention. Corpus-based descriptive research on the syntactic behaviour of the SCs has been a fairly recent endeavour. Siewierska et al. (2010) investigate the grammatical properties of SCs in the one-million-word Lancaster – Los Angeles Corpus of Spoken Chinese (LLSCC; Xiao and Tao 2006) and the one million-word Lancaster Corpus of Mandarin Chinese (LCMC; McEnery and Xiao 2004). They probe into the typical lexical and grammatical structures of split SCs and their frequencies in different genres, finding that the types of inserted elements in split SCs are limited in number and show a general structural pattern such as SC_H + NEG + ASP/RVC + MC + CL + MOD + SC_T.³ Their results also show that SCs appear much more frequently in such informal genres as fiction, stories, and conversations, as compared to general prose or academic writing. These observations are confirmed by Wang (2011), who studies the syntactic features of SCs in a much larger corpus of Modern Chinese, namely the CCL corpus developed at Peking University.⁴

Corpus-based translation studies and contrastive linguistics have been gaining momentum in China since a decade or so ago (cf. Wang and Huang 2008). One of the reasons might be the availability of some large Chinese-English parallel and comparable corpora. The bulk of the Chinese-English parallel corpus based studies in China have so far focused on the general distribution of lexical items. Syntactic structures like causation (via *ba*-construction) and passivity (via *bei*-

³ The following grammatical glosses are used in this article: ASP for aspect or aspect marker, BEI for the passive marker *bei*, CL for classifier, DE for structural particle *de*, GUO for the experiential aspect marker *-guo*, MC for numeral, MOD for modifier, NEG for negator, PRT for particle, QILAI for the inceptive aspect marker *-qilai*, RVC for resultative verb complement, and ZHE for the durative aspect mark *-zhe*.

⁴ The corpus is The Centre for Chinese Linguistics Corpus (<http://ccl.pku.edu.cn/corpus.asp>) developed at Peking University. The Modern Chinese component of the corpus contains 477 million Chinese characters.

construction) are also reduced to lexically based queries and analyses. Few studies have addressed the translation of particular types of syntactic or morpho-syntactic structures.

In our literature search, no Chinese-English parallel corpus based study on SCs has been identified. It is hoped that our research on the SCs and their English translations based on parallel corpus data will cast new light on the understanding of SCs in particular and discontinuous Chinese constituents in general. We aim to address the following two research questions in our study:

1. What are the patterns of non-correspondences between SCs and their English translations in terms of lexico-grammatical and semantic properties?
2. How can the non-correspondences between SCs and their English translations be accounted for from cross-linguistic and translational perspectives?

3 Methodology

3.1 Corpus

The parallel corpus used in our study consists of five million Chinese characters and English words in approximately 120,000 English-Chinese sentence pairs.⁵ The corpus includes both directions of translation and covers both fictional and non-fictional texts in four broad genres, namely fiction (60,622 sentence pairs), humanities (22,031 sentence pairs), social sciences (37,568 sentence pairs), and natural sciences (2,011 sentence pairs). All bilingual texts are sentence aligned and manually checked (cf. Wang 2004).

3.2 Retrieval of SCs and the processing of parallel concordances

The SCs investigated in the present study are the intersection of the most frequent SCs identified in two previous studies (Siewierska et al. 2010; Wang 2011). Siewierska et al. (2010) found 166 frequent SCs, based on the list in Yang et al. (1995) that contains 1,738 common SCs, in the Lancaster – Los Angeles Corpus of Spoken Chinese (LLSCC) for spoken Chinese and the Lancaster Corpus of Mandarin Chinese (LCMC) for written Chinese. Wang (2011) identified 60 typical SCs out of

⁵ The corpus size for the English component is computed by words, and the Chinese corpus size is measured by characters, the least ambiguous orthographic unit in Chinese running texts.

392 SCs in Peking University's CCL corpus. Our study regards the 22 common SCs shared by both studies as the most frequent SCs in Chinese. The 22 SCs are further divided into three groups, i.e. explicit verb-object group, pseudo-verb-object group, and idiomatic group (see Section 5.1), according to their internal semantic structures. Seven SCs are analysed as representatives of these three groups.

Each of the 22 SCs was searched and all instances of their split and non-split uses were retrieved from the parallel corpus using the parallel concordancer developed at Beijing Foreign Studies University (Wang and Xiong 2011). The tool returned all the parallel concordance lines containing the 22 SCs in both translation directions – Chinese-to-English and English-to-Chinese. Only Chinese-to-English translation results were included for further structural and semantic analyses, as the main focus of the present study is on Chinese-to-English translation.

The results were then exported into two text files, one for the split uses of SCs and the other for the non-split ones. The texts were manually checked for the valid cases of SCs. The structural constituency and semantic properties of the SCs were manually annotated.

Our data analysis consists of two phases. The first phase deals with the structural non-correspondences between 22 SCs and their English translations. The second phase is concerned with detailed qualitative and quantitative analysis of the argument structures in cases of non-correspondences of the seven representative SCs. In summary, the analyses focus on the translational non-correspondences and compare the preferred translation of non-correspondences between split and non-split usages of SCs in Chinese-to-English translation of SCs.

4 Morpho-syntactic patterns of SCs in Chinese-to-English translation

4.1 The extent of splittability of the SCs in native and translated Chinese

SCs are frequent and productive in Chinese, and both non-split and split uses of SCs are common verbal forms. Table 1 summarises the extent of splittability of the 22 frequent SCs in the parallel corpus. The data are broken down according to their usages, split (S) or non-split (NS). The ratios of separation, or splittability, are calculated using the formula: $\text{ratio} = S / (NS + S)$, where S and NS are respectively the frequency of split and non-split usage. As can be seen from Table 1,

Table 1: Non-split and split uses of SCs

SCs	Non-split	Split	Separation ratio
握手 <i>woshou</i>	2	9	82%
吸烟 <i>xiyan</i>	13	15	54%
洗澡 <i>xizao</i>	9	9	50%
翻身 <i>fanshen</i>	10	6	38%
睡觉 <i>shuijiao</i>	33	18	35%
告状 <i>gaozhuang</i>	2	1	33%
丢人 <i>diuren</i>	5	2	29%
操心 <i>caoxin</i>	3	1	25%
享福 <i>xiangfu</i>	3	1	25%
泄气 <i>xieqi</i>	6	2	25%
念书 <i>nianshu</i>	13	4	24%
鞠躬 <i>jugong</i>	7	2	22%
帮忙 <i>bangmang</i>	28	6	18%
吵架 <i>chaojia</i>	14	2	13%
理发 <i>lifa</i>	7	1	13%
犯罪 <i>fanzui</i>	92	12	12%
搞鬼 <i>gaogui</i>	1	0	0%
鼓掌 <i>guzhang</i>	8	0	0%
化妆 <i>huazhuang</i>	2	0	0%
敬礼 <i>jingli</i>	1	0	0%
伤心 <i>shangxin</i>	36	0	0%
跳舞 <i>tiaowu</i>	3	0	0%
Total	298	91	

some SCs are more likely to be found in split forms than others, for example, 握手 *woshou* ‘shake hands’ (82%), 吸烟 *xiyan* ‘smoke’ (54%), 洗澡 *xizao* ‘take a shower’ (50%) and 翻身 *fanshen* ‘turn over, or figuratively, to become rags-to-riches type of successful’ (38%) are often used in a discontinuous manner.

4.2 Patterns of morpho-syntactic non-correspondences between SCs and their English translations

In our data, non-split SCs, predominantly functioning as verbs, are often found to be translated into English as a variety of word classes, such as verbs, nouns, adjectives, and adverbs. Table 2 shows the major English equivalents of the 22 SCs in Chinese source texts in the descending order of their frequencies. The term ‘English equivalents’ means that the English expressions are to a large extent

Table 2: Typical English equivalents of the 22 SCs

犯罪 <i>fanzui</i>	crime(s); (non)criminal(s); sin, sinned, sinner; offence, offend
睡觉 <i>shuìjiào</i>	sleep, sleeping, asleep, slept, sleeper, sleeplessness; go/get/ to bed
伤心 <i>shāngxīn</i>	sad, sadly, saddened, sadder; hurt, be hurt, hurting; grief, grieves, grieved, grieving; mournful; heartbreak, heartbreaks, heartbroken, heartbreaking, break one's heart, heart was broken
帮忙 <i>bāngmáng</i>	help, helped, helper, helpful(ly); assistance, assist; aid, unaided
吸烟 <i>xīyān</i>	smoke, smoking, smokers, non(-)smokers, smoked
洗澡 <i>xǐzǎo</i>	bath(ing); wash(ing)
念书 <i>niànshū</i>	study
翻身 <i>fānshēn</i>	turn/roll over
吵架 <i>chǎojiǎ</i>	brawl(ing); fight(s)
握手 <i>wòshǒu</i>	shake, shaking one's hands
鞠躬 <i>jūgōng</i>	bow, bowed, bowing
泄气 <i>xièqì</i>	(dis)encouragement; dismayed
理发 <i>lǐfā</i>	barber
鼓掌 <i>gǔzhāng</i>	applause; clap, clapper
丢人 <i>diūren</i>	mean; shame
操心 <i>cāoxīn</i>	bother(ed); worry, worried; trouble, troubled; care, cared
享福 <i>xiǎngfú</i>	N/A
告状 <i>gào zhuàng</i>	tell
跳舞 <i>tiàowǔ</i>	dance, danceable, danced, dances, dancing; ball
化妆 <i>huàzhuāng</i>	toilet goods
搞鬼 <i>gǎoguǐ</i>	N/A
敬礼 <i>jìnglǐ</i>	salute, saluting

‘faithful’ to the original SCs. Liberal translation, however, is also possible, for example, in the case that the meaning of an SC is culturally rich. For instance, the SC 翻身 *fānshēn* is translated into ‘get rid of poverty’ in example (1) below.⁶ In some other cases, no word or phrase can be identified as the equivalent of an SC. This is often because the context explains its meaning.

(1) 有的地方一年翻身，有的地方两年翻身。

youde difang yi nian fānshēn, youde difang liang nian fānshēn
 some place one year turn-over some place two year turn-over
 ‘In some places, it took only one or two years to get rid of poverty.’

⁶ Radical change to the sense of the verb often causes great transformation in the argument structure, which undermines the common ground for comparison.

The split uses of many SCs have more than one pattern of correspondence in English translations. For instance, 犯罪 *fanzui* can be translated into the verbal phrase ‘commit crime’ or the noun or adjective ‘criminal’. Likewise, 翻身 *fanshen* is sometimes translated literally into ‘turn over’ and sometimes metaphorically as ‘emancipate’.

Three major types of correspondences are observed for the non-split SC equivalents: 1) the first type is composed of a light verb (written as the small letter *v* henceforth) and a nominal phrase (NP). 洗澡 *xizao* is such a case in point, which is often translated into ‘take/have a bath’; 2) the second type consists of a lexical verb and its complement(s). 翻身 *fanshen*, for example, often corresponds to ‘turn over’; 3) in the third type, the meaning of an SC is almost evenly shouldered by the verb and the trailing nominal element or complement. In the case of 犯罪 *fanzui*, for example, the structure ‘commit ... crime’ is the predominant translation pattern.

Unlike split SCs, the patterns of correspondence of the non-split SCs are predictable and regular in the sense that there is a relatively limited inventory of corresponding verbs or phrasal verbs in English. The non-split SCs are more likely to correspond to single words or simple structures (such as *V + N*).

The patterns of correspondence in split SCs, however, deserve more attention. The phrase 搞什么鬼 *gao shenme gui* is translated into (part of) a clause *What the devil do you do*. The interposing question word 什么 *shenme* ‘what’ serves as the complementiser of an interrogative clause in English. When a split SC corresponds to the structure ‘*v + NP*’, the makeup of the NP might be rather complex. For example, the split SC 鞠了一躬 *ju-le yi gong* is translated rather differently at an initial glance, e.g. ‘made a neat bow’, ‘made her a bow of exaggerated politeness’, ‘made her a burlesque bow’, ‘made her another bow’, ‘made his bow’, ‘made his formal bow’, ‘making a slight bow’, ‘making rather a pert bow’, and ‘with a grave obeisance’. However, most of them do share the same verbal frame, namely, the light verb ‘make’ plus the noun ‘bow’. Of great interest to the different translations of 鞠了一躬 *ju-le yi gong* lies in the relocation of preverbal modification (as in 浅浅地鞠了一躬 *qianqian de ju-le yi gong* ‘slightly bowed’) to the pre-modifier of the nominalised ‘bow’, that is, the manner of the action is glued to the SC as an inseparable constituent. Additionally, in the cases ‘made his bow’ and ‘made her another bow’, the agent and the patient are integrated in the SC construction respectively.

Motivated by such morpho-syntactic variability of SC translations, we find it particularly interesting to make in-depth analyses of the argument structure change of the SC frames such as the addition, transposition, and alteration of semantic roles.

4.3 Interim summary

On the basis of our corpus analysis, considerable differences have been identified between the lexico-grammatical patterns of English translations of non-split and split SCs: the former tends to be translated into single words or simple phrases, and the split ones into complex verbal structures; the non-split SCs have a limited variety of frequent English translation patterns, which is not the case in the English translations of split SCs. The English correspondences of non-split SCs are regular and stable whereas the correspondences of split SCs are notably more diversified given different sentential contexts.

The English translations of the split and non-split uses of the same SC often share similar verbal frames (i.e. contiguous and non-contiguous constructions containing verbs and their objects). Likewise, the translations of short split SCs, most probably, tri-syllabic ones, often resemble those of the non-split SCs. For example, in short split SCs of 帮...忙 *bang...mang*, such as 帮过忙 *bang-guo mang*, 帮了忙 *bang-le mang*, or even 帮他一点儿忙 *bang ta yidianr mang*, most English translations are ‘help’. Therefore, the extent of structural non-correspondences in English correlates with the splittability of SCs, with the non-split SCs at one end, and the most complicated split SCs at the other.

The next section will investigate the extent to which the semantic properties of SCs vary between SCs and their English translations via case studies of the seven representative SCs.

5 Semantic patterns of SCs in Chinese-English translation

5.1 Overall semantic non-correspondences of seven SCs

Tertium comparationis in a cross-linguistic study has to be independent of formal features. In our study, verb senses and argument structure have been pinpointed as such a basis for comparison. To be fair to both languages, we use the common core semantic roles (i.e. Agent, Dative, Experiencer, Manner, Patient, and Theme) outlined in Riemer (2010) and Yuan (2010), among the very many different versions of argument structure theories.

In the analysis hereafter, seven SCs are chosen from the 22 most frequent SCs, as they are representative of three different grammatical structures: 握手 *woshou* ‘shake hands’, 吸烟 *xiyan* ‘smoke (cigarettes)’, and 念书 *nianshu* ‘read books; study’ form Group A as the explicit verb-object group; 帮忙 *bangmang* ‘help’,

Table 3: Morpho-syntactic and semantic non-correspondences of SCs and their translations

Degree of change	Dimension of change	Tags	Explanation of tags
Radical changes	Verb senses	semantic change	The meaning of SC is changed.
		-v_sense	Part of verb meaning is dropped.
		-pred.	The SC is replaced by a prepositional phrase or otherwise.
Non-radical changes	Argument structure	ellipsis	complete omission
		+argu.	argument addition
		argu.alternation	argument alternation (e.g. agent in SC becoming patient in English)
		-argu.	argument omission
		argu.incorporation	argument incorporation (e.g. two semantic roles in SC becoming one in English)
		argu.split	split argument (e.g. one semantic role in SC becoming two in English)
	Lexico-grammar	grammar	grammatical change
		sequence	word order change
		lexicon	lexical change

鞠躬 *jugong* ‘bow’, and 洗澡 *xizao* ‘bathe’ form Group B as the pseudo-verb-object group; and Group C is the most idiomatic group, with one member 操心 *caoxin* ‘care’, the initial and the final morpheme of which are grammatically unanalysable.

The non-correspondences between SCs and their English translations are scrutinised and carefully annotated in the retrieved sentence pairs. Table 3 tabulates the categories of non-correspondences according to their argument structure as well as lexico-grammatical properties. The adjustments in argument structure in the English translations of SCs are further divided into five types of non-correspondence in semantic roles (argument addition, argument alternation, argument omission, argument incorporation, and split argument, see Section 5.2.1). Three types of change in the lexico-grammatical realisations (grammar, sequence and lexicon, see Section 5.2.2) in English translations are annotated; and radical treatments or free translations of SCs are placed under the name of ‘radical changes in verb sense’. Radical semantic changes, which involve changed meaning, are rare and they somehow go to the creative, artistic or sometimes erroneous, side of translation. Radical changes are hardly generalisable despite the availability of statistical data for changes of this nature. Hence the following discussion will focus on non-radical changes, though data of radical changes will be cited where appropriate and necessary.

Table 4: Normalised frequencies of non-correspondences

Dimension of change	Categories	握手 wo shou	吸烟 xi yan	念书 nian shu	操心 cao xin	帮忙 bang mang	洗澡 xi zao	鞠躬 ju gong
Verb senses (non-radical changes)	semantic change	0	36	529	0	29	0	0
	ellipsis	0	0	59	0	88	56	111
	–pred.	91	179	353	0	118	167	333
	–v_sense	0	0	0	0	59	0	0
Semantic role	+argu.	273	107	0	1000	235	389	778
	argu.	0	0	0	250	147	0	0
	alternation							
	–argu.	0	393	59	250	29	0	0
	argu.	0	0	0	0	0	0	0
	incorporation							
Lexico-grammar	argu.split	0	0	0	250	59	0	0
	grammar	0	107	0	1000	265	111	222
	sequence	182	0	0	750	265	111	222
	lexicon	91	0	0	250	0	0	0

Table 5: Percentage of the grouped non-correspondences for seven SCs

Dimension of change	握手 wo shou	吸烟 xi yan	念书 nian shu	操心 cao xin	帮忙 bang mang	洗澡 xi zao	鞠躬 ju gong
Verb senses (non-radical changes)	9%	21%	94%	0%	29%	22%	44%
Semantic roles	0%	39%	6%	75%	24%	0%	0%
Lexico-grammar	27%	11%	0%	200% ⁷	53%	22%	44%

Table 4 shows the normalised frequency counts (per 1,000 non-correspondences) of the seven SCs in different types of non-correspondences. The aggregated proportions of different non-correspondences are then reported as three broad categories in Table 5. The percentage data indicate that verb-object SCs such as 念书 *nianshu* ‘read books; study’, 帮忙 *bangmang* ‘help’, and 吸烟 *xiyan* ‘smoke (cigarettes)’ show a high propensity for non-correspondences such as semantic change, or -pred (e.g. omission of lexical verb) in translation.

7 In all the translations of *caoxin*, both grammar and sequence change were spotted and tallied, thereby adding up to 200%.

Pseudo-verb-object SCs such as 帮忙 *bangmang* ‘help’, 洗澡 *xizao* ‘bathe’, and 鞠躬 *jugong* ‘bow’ are more likely to be subject to change in the argument structure.

The next section will present a detailed analysis of the major types of semantic non-correspondences in the English translations of Chinese SCs.

5.2 Semantic non-correspondences between SCs and their English translations

5.2.1 Non-correspondences in the argument structure

Non-correspondences in the argument structure refer to the misalignment in argument structure between an SC and its English equivalent. It is, however, worth noting that these labels are only used on those less radical non-correspondences where verb senses largely remain intact. A total of five types of common semantic non-correspondences are identified, which include the addition and deletion of an argument in English equivalents (marked as +argu and –argu respectively in Table 4), the incorporation and the splitting of arguments (argu_incorporation and argu_split), and exchange in the semantic role of a particular NP (argu.alter-nation). The most frequent semantic non-correspondences, namely argument addition, omission, and alteration, are illustrated with examples as follows.

5.2.1.1 Argument addition

Table 6 summarises the frequency of each type of added argument in the English translations of SCs. Agent, manner, patient, theme and dative are likely to be added in translation.

Examples (3) and (4) illustrate the cases where the key arguments are absent in Chinese sentences. In example (3), the subject or the agent of 握手 *woshou*

Table 6: Frequency of changed arguments in argument addition

Argument	Frequency
Agent	8
Manner	7
Patient	4
Theme	4
Dative	3
Experiencer	1

‘shake hands’ is missing. Native Chinese readers, however, can figure it out effortlessly by referring to the context.

- (3) 说完紧握着他的手。
shuo wan **jin** **wo-zhe** **tade shou**
say finish tightly hold-ZHE his hand
‘At this *she* **squeezed his hand.**’
- (4) 我承刘东方帮过忙，……。
wo cheng Liu Dongfang **bang-guo mang**
I be-indebted Liu Dongfang help-GUO busy
‘I’m indebted to Liu Tung-fang for **having helped me once.**’

English, on the other hand, does not seem to encourage omissions in such circumstances and the English counterparts of the sentences specify all the missing arguments in order to guarantee the semantic integrity as well as the rule of English grammar. As a result, in example (3) ‘*She* squeezed his hand’, the agent *she* is added; likewise, the patient is spelt out in example (4) ‘I’m indebted to Liu Tung-fang for having helped *me* once’. This points to the typological difference between the two languages when dealing with the supplemented arguments (either previously mentioned or readily accessible from the context). The Chinese argument system relies more on the context, whereas English requires its lexico-grammar at full array. Such a cross-linguistic difference, therefore, tends to encourage the translator to specify arguments when translating from Chinese to English.

5.2.1.2 Argument omission

In our data, the incidence of argument omission in English is close to that of argument addition. The most frequently omitted argument in English is the patient (see Table 7).

Table 7: Frequency of changed arguments in argument omission

Argument	Frequency
Patient	11
Agent	1
Experiencer	1
Manner	1
Theme	1

- (5) 我从口袋里掏出野梨花木烟斗，拧上一斗烟丝，坐在地下吸起烟来。

wo cong koudai li tao chu ye lihuamu yandou, ning shang yi dou
 I from pocket in take out wild pear-wood pipe roll on one pipe
 yansi, zuo zai dixia xi qi- yan -lai
 tobacco, sit on ground smoke QI- tobacco -LAI
 'I pulled my pear wood pipe out of my pocket, rolled a plug of tobacco, and
 sat down to **smoke**.'

In example (5), the SC 吸烟 *xiyan* 'smoke (cigarette or tobacco)' is translated into the verb 'to smoke', which is often used intransitively in English, though 'smoke a cigarette' is not strictly impossible. Thus, the nominal tail of the SC is omitted when translated into English. The elaboration of the 'cigarette' can be regarded as a form of periphrasis. By stretching the SC into a verbal complex, such periphrasis renders it more legitimacy of syntactic independence, so that it can function as a clause by itself.

5.2.1.3 Argument alternation

Argument alternation occurs in the case where the same nominal phrase has different semantic roles in a Chinese sentence and its English translation. The majority of such non-correspondences are found on datives of SCs, which often correspond to the patient of the English verb. In example (6), the nominal phrase 地方 *difang* 'civilian units' acts as the object of the preposition 对 *dui* 'toward, to' before the SC.

- (6) 这也是我们军队自己的事情，所以军队对地方要支持，要帮忙。

zhe ye shi women jundui ziji de shiqing, suoyi jundui dui difang
 this also is we army self DE affair, so army to civilian-unit
 yao zhichi, yao **bangmang**
 should support should help
 'But this also concerns the army, so it should support and **help** the civilian
 units involved.'

According to Yuan (2010: 134), the nominal phrase 地方 *difang* 'civilian units' in example (6) is analysed as the dative of 帮忙 *bangmang* 'help'. In contrast, its English counterpart, 'the civilian units', functions as the patient rather than dative of the verb 'help'. Such an alternation can affect the semantic features of the verbal complex in a subtle manner. To put it differently, while the nominal phrase in Chinese is only used to specify the recipient of the action, its English counterpart bears the extra meaning of 'affectedness'. The frequent alternations

between Chinese datives and English patients are associated with another typological difference in the syntax between the two languages. In Chinese syntax, a slot for prepositional phrases is expected in the immediate left context of some verbs (e.g. 敬礼 *jingli* ‘salute’, 握手 *woshou* ‘shake hands’). The prepositional phrases are usually led by 对 *dui* ‘for (somebody)’ or 与 *yu* ‘with (somebody)’ that introduces the recipient of the action (compare ‘for somebody’ in English). English, however, does not allow the prepositional phrase to be placed immediately to the left of the verb so that forms like ‘*for somebody help’ are ungrammatical. Moreover, there seems to be a difference in the syntactic behaviour of the verbs themselves in the two languages. While the English verb ‘help’ can be used transitively and take an object, its Chinese counterpart 帮忙 *bangmang* ‘help’ can only be used intransitively in its non-split form. Forms like *帮忙他 *bangmang ta* are ungrammatical in Chinese. In a nutshell, the majority of Chinese SCs allow a pre-verb dative but not a patient whereas English verbs allow a patient but not a pre-verb dative. Such cross-linguistic differences are potential sources of difficulty for translators. This illustrates the complicated inter-relationships between form and meaning in translation, whose interactions are discussed in the next section.

5.2.2 Morpho-syntactic non-correspondences between SCs and their English translations

In the translation practice between Chinese and English, translators sometimes have to adjust the morpho-syntactic realisations of a unit of meaning. This has manifested itself in our earlier analysis of the changes in argument structure of SCs and their English equivalents. The most salient patterns of change in grammatical property and word order are summarised in Table 8.

Table 8: Changes in grammatical properties

Grammatical property in Chinese	Grammatical property in English	Frequency
prep._obj.	obj.	5
obj.	prep._obj.	4
prep._obj.	direct_obj.	2
obj.	obj.	1
sub.	prep._obj.	1
sub.	direct_obj.	1
prep._obj.	sub.	1
obj.	sub.	1
obj.	adjunct	1
direct_obj.	indirect_obj.	1

Grammatical property alternation seems to centre on a few grammatical categories: the prepositional object, indirect and direct object and adjunct. The most frequent pattern is the transformation between prepositional objects and verbal objects. Prepositional objects are frequently transferred into verbal objects (both direct and indirect objects) in the translation between Chinese and English. The transformation between the prepositional object and the verbal object is illustrated in example (7). The Chinese prepositional object 我们 *women* ‘we, us’ corresponds to the object of English verb ‘bathe’, namely ‘us’.

- (7) 天色大明了。两个看护脸上堆着新年的笑，走了进来，替我们洗了澡。
 tianse da ming-le. liang ge kanhu lian shang dui-zhe xinnian
 sky-colour broad bright-PRT, two CL nurse face on pile-ZHE new-year
 de xiao, zou-le jinlai, ti women xi-le zao.
 DE smile, walk-LE come-in, for us wash-LE bath
 ‘In the morning two nurses beaming with New Year smiles came to **bathe** us.’

Similarly, Chinese objects also correspond to English prepositional objects, which are illustrated in example (8). It can be argued that the transformation of both directions (verbal object to prepositional object and the opposite) is the result of the lexico-grammatical differences between Chinese and English. English does not use a pre-verb prepositional phrase which marks the verb’s dative. English has some grammatical structures which Chinese does not share either. To summarise, grammatical changes are often concomitant with semantic structure changes in Chinese to English translation.

- (8) 越想越恨，泪被怒火截住，他狠狠地吸那支烟，越不爱吸越偏要吸。
 yue xiang yue hen, lei bei nuhuo jiezhu, ta henhende xi
 the-more think the-more hate, tear BEI anger stop he firmly smoke
 na zhi yan, yue bu ai xi yue pian yao xi.
 that CL cigarette, the-more not like smoke the-more insistently want smoke
 ‘Bitterness flooded his soul, blazing anger dried his tears and he **puffed**
furiously at his cigarette. The more he disliked smoking, the harder he
 smoked.’

So far, we have had a detailed analysis of the most typical structural and semantic non-correspondences between the seven representative SCs and their English translations. Salient patterns are summarised in each of these groups with statistical evidence. The discussion on this topic is continued in the next section, which will explore the influence on the patterns of the non-correspondences motivated by the inherent grammatical structure of the SCs.

5.3 Influence of SC splittability on non-correspondences

In this section, comparisons are made between the non-correspondences of split and non-split SCs in Chinese to English translation. Table 9 compares each of the non-correspondence categories, structural or semantic.⁸ Among the 12 categories, argument addition (+argu), semantic change (semantic alternative), argument alternation (argu.alternation), and argument omission (–argu) occur more frequently in the English translations of non-split SCs.

In the data with the non-correspondence of argument addition, there is a general tendency to leave out the patient in Chinese (hence English clauses acquire extra arguments). All of the patient omissions appear in the translation of 帮忙 *bangmang* ‘help’, which is intransitive and can take the recipient, for example, by the pre-verb prepositional phrase 为 ... 帮忙 *wei ... bangmang* ‘for (someone) help’. Thus the English verbal phrase receives one more patient than its Chinese counterpart does. The split use of 帮忙 *bangmang* ‘help’ in Chinese, however, prepares an extra slot for the recipient. The expanded compound allows the recipient to be incorporated as the modifier of SC_T, which means that 为 ... 帮忙 *wei ... bangmang* can be transferred into 帮 ... 的忙 *bang ... de mang*. As a result, the split SC bears more resemblance in its argument structure to its English equivalent than its non-split counterpart.

The extra grammatical slot of split SCs may also account for the relative unpopularity of argument alternation in split correspondences. Among the pat-

Table 9: Comparison of non-correspondences in split and non-split SCs

Categories	NS	S
+argu.	23	4
semantic alternative	11	1
argu.alternation	6	0
–argu.	10	5
ellipsis	4	1
-v_sense	2	0
–pred.	10	9
grammar	6	12
sequence	8	9
lexicon	0	2
argu.split	0	3
argu.incorporation	0	0

⁸ Given that the values in many of the cells in Table 9 are less than 5 or even zero, no statistical test has been adopted.

terns of argument alternations in the English translations of non-split SCs, the most significant one is the correspondence between Chinese dative and English patient, i.e. the cases where the prepositional phrase ‘为 *wei* ‘for’ (or 给 *gei* ‘for’) + dative + SC’ in Chinese corresponds to ‘V + patient’ in English. Such a pattern, nevertheless, is made less prevalent by the extra slot of modifier. The argument alternation is therefore less frequently used in translating split SCs.

More non-correspondences in grammatical properties may be the result of the more complicated internal structure of the expanded compounds, which have extra spaces for elements such as aspect markers, participants and manner while keeping all of the external slots. This offers more choices in translation and makes grammatical changes more complicated.

A large proportion of argument addition occurs on the subject and object of the verb phrase (16 out of 27). The subjects and objects are omitted in Chinese, because Chinese allows the absence of both, as they are most likely to be inferred from the context. Such dropping of subjects and objects is impermissible in English, however. Thus the addition of these arguments constitutes a large portion of argument addition in Chinese-to-English translation of SCs.

6 Discussion

The translational non-correspondences can be explained by the differences in morpho-syntax between Chinese and English, two typologically distant languages. Such differences can sometimes confuse bilingual language users and translators when they deal with linguistic forms such as SCs. In the Chinese linguistic system as well as Chinese orthography, 汉字 *hanzi* ‘Chinese character’ is probably the least ambiguous linguistic unit; however, the boundaries of word and/or phrase are far less well-defined. What upsets a Chinese-to-English translator is that an English word normally corresponds to more than one character in Chinese. For instance, the Chinese character 吸 *xi* ‘inhale, absorb’ alone does not bear the meaning of ‘smoke’, because it can combine with various other characters (e.g. 吸水 *xishui* ‘absorb water’, and 吸热 *xire* ‘absorb heat’). Thus, the addition of an object is required when the English word ‘smoke’ is translated into Chinese as 吸烟 *xiyan* ‘smoke (cigarette or tobacco)’ whereas the explicit object is often unspecified when the Chinese word 吸烟 *xiyan* is translated into English. Moreover, Chinese multi-character units are less stable than English words and/or compounds, and their boundaries with phrases are fuzzy. SCs are a typical case. However, English does not allow verbs to be disjoined in halves and interrupted with other lexico-grammatical elements. Such differences are at least part

of the difficulty of finding consistent patterns of English translation for split and non-split SCs.

Linguists' interests in SCs are due to the flexibility of the structural and semantic properties when SCs are used in their split or non-split forms. The insertion of grammatical elements in SCs turns the split SCs into verb-object phrases, while their non-split counterparts are used as compound verbs. The extension of the SC morpho-syntax with insertions, as a result, increases the semantic capacity of the compounds, and the inserted elements tend to be integrated into the semantic structure of the whole verb. This also accounts for the incorporation of participants and manner of actions in split SCs, which is semantically realised by the change in semantic roles or the change in the overall argument structure in pseudo verb-object SCs. The addition of linguistic elements turns the dummy object into a grammatically necessary one, specifying the participant or the manner of the action.

The differences in the lexico-grammar and semantics of split and non-split SCs are inherited by or mapped onto their English translations, but not without variability. Our general observation is that typical English equivalents of split uses of SCs are verbal phrases, longer and more complex than their non-split counterparts, which are usually translated as verbs or simple phrases. The translations of non-split SCs have a limited inventory of English verbs and verbal phrases, but the translations of split SCs are characterised by more diversified English verb types and structures. Also, the differences in grammatical structures are sensitive to the differences in semantic structures. There is a tendency for the English translation of split SCs to have more than one semantic focus (e.g. the verbs and their objects), and the focus of the expanded SCs often leans towards the verb satellites (object, pre- and post-verbal elements), rather than the verbs themselves. In non-split cases, however, the verbs are very often the semantic foci.

The predominant pattern revealed by our corpus investigation is that the English translations of SCs are characterised by semantic explicitation as well as structural explicitation (cf. Baker 1993, 1996). The English translations of split and non-split SCs behave rather differently in terms of their lexico-grammatical and semantic structures. The translations of non-split SCs more often involve rearrangements of argument structure, such as argument addition and semantic alternative, while changes in grammatical properties are more often involved in translating split SCs.

The SCs are conceived of as the middle ground between words and phrases in Chinese, and the difference also figures prominently between Chinese and English morpho-syntactic systems. Differences in phraseology have also been identified in the constituency of Chinese SCs and their equivalents in English.

Some SCs have structural constraints on the inserted elements. For example, only certain elements can be inserted as pre-modifiers of SC₁s. Their English equivalents, however, are not subject to such constraints. The differences often trigger the rearrangement of relevant elements. In the translation between the SC 鞠躬 *jugong* and ‘make + a + adj. + bow’, for example, the modifier of ‘bow’ is preposed and/or becomes the modifier of the whole SC.

In many cases, the translation of SCs is not only affected by their internal structures, but also by their syntactic contexts. The difference in grammatical constraints at the clause level is often responsible for translational non-correspondences. The following syntactic differences have been identified in the translational non-correspondences of SCs: first of all, in Chinese, the subject and the object are optional, which contrasts with English. When Chinese is translated into English, the implicit subjects are recovered. Secondly, Chinese and English differ in the ordering of syntactic elements. In Chinese clauses, modifiers typically precede verbs, whereas in English they tend to follow verbs. Changes in grammatical status or word order are therefore not uncommon when translating such phrases, and as a result the semantic role of the nominal phrase changes accordingly (e.g. from the dative in Chinese to the patient in English and vice versa). Thirdly, the clausal and sentential structures of Chinese and English are also different. The constituent structure of English clauses and sentences are clearly marked: every sentence takes only one main clause, which typically bears one main verb. The remaining elements depend upon and are attached to the verb. Chinese sentences, however, are more coordinative than hierarchical, or regressive in generative grammar terms. This is pertinent to the arguably notorious typological distinction of hypostasis versus parataxis between English and Chinese. Chinese sentences are often the juxtaposition of small clauses, or phrases or even words, and it is not easy to identify the main clause on which all other clauses depend. In each of these clauses, the verbs are also frequently found to appear in serial constructions without explicit grammatical hierarchy. Thus a complex adverbial phrase in an English sentence may correspond to an SC and gain syntactic independency. Therefore, some verbs are omitted or made syntactically dependent when a Chinese verb is translated into a noun in English.

The simple and straightforward goal of translation is to translate the meaning of the original text into another language. Meaning is therefore the common ground for the source text and the translated text. Yet this goal is hardly easy, if not impossible, to achieve in translation practice, not because a piece of message cannot be articulated with another set of linguistic forms, but because of the non-correspondences of argument structure between two languages.

In our examination of the English translations of SCs, we are aware that the sentences in translated texts are more semantically explicit than their counter-

parts in original Chinese. In original texts, the meaning is expressed in a more opaque manner and context dependent. A word tends to convey a more generalised sense (e.g. 'do something' expresses the idea of 'giving aid'), and the inferable arguments are often omitted (e.g. the subject dropping in Chinese). In English translated texts, however, the meaning is spelt out in greater detail, or is more semantically explicit. Verbs with a general sense in Chinese are substituted by more specific expressions in English translations; and covert arguments in Chinese are made more explicit in English translations. Sometimes, words and phrases are added in English translations to explain the manner and purpose of an action. Semantically alternative terms are sometimes preferred over the less lucid expressions in English.

The split and non-split uses of SCs have different grammatical and semantic structures, which leads to differences in grammatical and semantic structures of their English translations. The overarching contributing factor behind such operations relates to the typological differences between English and Chinese. Above all, the differences or non-correspondences in the translations of split and non-split SCs reveal the primary purpose of translation, that is, to recode the conceptual and/or semantic substance of the original text in another language.

7 Concluding remarks

This study compares the morpho-syntactic and semantic properties of the split and non-split uses of SCs in Chinese and their English translations. It starts with the descriptive analysis of the lexico-grammatical structures of split/non-split uses of SCs and their English translations. This is followed by a detailed analysis of seven typical SCs and their English equivalents with special attention to the semantic non-correspondences. With evidence from the Chinese-English parallel corpus, patterns of translational non-correspondences have been identified in the light of the grammatical and semantic structures. We conclude that non-split SCs tend to be translated into words or simple phrases in English, whereas split uses are often translated into verbal complexes which sometimes have rather different morpho-syntactic choices.

This research is one of the few corpus-based systematic investigations into the English translations of Chinese SCs, which offers new insights into the comparison and contrast of the two language systems in lexico-grammatical and semantic dimensions. Another objective, less explicitly expressed though, is to propose a standard procedure for systemically evaluating bilingual equivalents in terms of their semantic-grammatical structures. This model incorporates both grammatical and semantic (i.e. argument structure) properties and can be

adopted in other similar cross-linguistic studies, for example, separable verbs in German, Dutch, Hungarian and Afrikaans, verb complement constructions in Chinese, and phrasal verbs in English.

Finally, we are aware that the factors discussed in this research are by no means the only ones that impact on the choice of predicates in translation. Evidence from corpora has also shown that other factors including discourse-pragmatic concerns such as genre variation and information packaging also play important roles in producing different translations, which deserve future research. It is hoped that this study will make theoretical and methodological contributions to the Chinese verbal complexes and have its due impact on the research of contrastive studies of linguistic phenomena of similar kind.

Acknowledgements

We thank the anonymous reviewers for their very constructive comments on our manuscript. This research is funded by China's National Social Sciences Foundation grant "Parallel Corpus Based Study of Chinese Verb Complexes and Their Translations" (grant ref. 12CYY060) and partially supported by China's National Social Sciences Foundation Key Project "The Construction and Processing of Large Scale Chinese-English Parallel Corpus" (grant ref. 10ZD&127) and the Programme for New Century Excellent Talents in University (grant ref. NCET-12-0790).

References

- Baker, Mona. 1993. Corpus Linguistics and Translation Studies: Implications and applications. In Mona Baker, Gill Francis & Elena Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). Amsterdam: John Benjamins.
- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.), *Terminology, LSP and Translation* (pp. 175–186). Amsterdam: John Benjamins.
- Feng, Zhiwei. 2006. Evolution and present situation of corpus research in China. *International Journal of Corpus Linguistics*, 11(2). 173–207.
- Lu, Zhiwei. 1957. *Hanyu de Goucifa* (Chinese Morphology). Beijing: The Scientific Press.
- McEnery, Tony and Xiao, Richard. 2004. The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In Rute Costa, Maria Teresa Lino, Fátima Ferreira and Raquel Silva (eds.), *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC) 2004* (pp. 1175–1178). Lisbon: Centro Cultural de Belem.
- Riemer, Nick. 2010. *Introducing Semantics*. Cambridge: Cambridge University Press.

- Siewierska, Anna, Xu, Jiajin and Xiao, Richard. 2010. *Bang-le yi ge da mang* (offered a big helping hand): A corpus study of the splittable compounds in spoken and written Chinese. *Language Sciences*, 32(4). 464–487.
- Wang, Haifeng. 2011. *Xiandai Hanyu Liheci Lixi Xingshi Gongneng Yanjiu* (The Forms and Functions of Split *Liheci* in Modern Chinese). Beijing: Peking University Press.
- Wang, Kefei. 2004. *Shuangyu Duiying Yuliaoku Yanzhi yu Yingyong* (Parallel Corpus: Construction and Application). Beijing: Foreign Language Teaching and Research Press.
- Wang, Kefei and Huang, Libo. 2008. Yuliaoku fanyixue shiwu nian (Corpus-based translation studies: Progress in recent 15 years). *Foreign Languages in China*, 6(6). 9–14.
- Wang, Kefei and Xiong, Wenxin. 2011. Hanying duiying yuliaoku de jiansuo yu yingyong (Design and application of sentence pair retrieval from parallel corpora for translation studies and translation teaching). *Computer-Assisted Foreign Language Education*, (6). 31–36.
- Xiao, Richard and Tao, Hongyin. 2006. The Lancaster – Los Angeles Spoken Chinese Corpus. Lancaster: UCREL, Lancaster University.
- Yang, Qinghui (ed.). 1995. *Xiandai Hanyu Liheci Yongfa Cidian* (A Dictionary of Splittable Compound Usage in Modern Chinese). Beijing: Beijing Normal University Press.
- Yuan, Yulin. 2010. *Hanyu Peijia Yufa Yanjiu* (A Valance Grammar of Chinese). Beijing: The Commercial Press.

Bionotes

Jiajin Xu is Associate Professor of Linguistics at the National Research Centre for Foreign Language Education, Beijing Foreign Studies University as well as a founding member of the Chinese Society of Corpus Linguistics (CLSC). He obtained his PhD, specialised in Corpus-based Discourse Studies, from Beijing Foreign Studies University, China. From 2008 to 2009, he was post-doctoral researcher in the Department of Linguistics and English Language at Lancaster University, UK, working on the ESRC funded project “A Corpus-based Study of Split Words in Chinese: Interfacing morphology, syntax and pragmatics”. He has published a number of English papers in international journals including *Language Sciences*, *Chinese Language and Discourse* and *ICAME Journal*, and over 20 papers on corpus-based discourse studies in top-ranking linguistics journal in China. His recent books include *The Use of Discourse Markers in Spoken Chinese of Urban Teenagers* (Beijing 2009) and *Using Corpora: A Practical Coursebook* (co-authored, Beijing 2010).

Xiaochen Li is a doctoral student at Beihang University, Beijing, China. His major research interests include corpus linguistics, contrastive linguistics, discourse studies and pragmatics.