

# Unit 22 Domains, text types, aspect marking and English-Chinese translation (Case study 6)

## 22.1 Introduction

This unit uses comparable and aligned parallel corpora (see unit 5) to approach contrastive and translation studies as discussed in units 10.6 and 15.2-15.3. In unit 10.6, we noted that while aligned parallel corpora are well suited for translation studies, they provide a poor basis for contrastive language study if used as the sole source of data. They should most often be used in conjunction with comparable L1 language corpora. This unit extends this argument via a case study of aspect marking in English and Chinese. But before presenting the case study in detail, it is appropriate to provide background knowledge of aspect markers in Chinese.

Both English and Chinese mark aspect grammatically, but the aspect systems in the two languages differ considerably. As far as grammatical aspect (also known as *viewpoint aspect* as opposed to *situation aspect*, see unit 10.9) is concerned, English marks the progressive, the perfect, the perfect progressive and the simple aspect (cf. Biber et al. 1999: 461). In contrast, Chinese is more aspectual in that it has four perfective viewpoints (marked by *-le*, *-guo*, verb reduplication and resultative verb complements RVCs) and four imperfective viewpoints (marked by *-zhe*, *zai*, *-qilai* and *-xiaqu*) in addition to a number of complex viewpoints (see unit 15.3 for a discussion of aspect marking in Chinese). Aspectual meanings in Chinese can be realized in three ways: (i) marked explicitly by aspect markers, for example *-le*, highlighted in (1a), (ii) marked adverbially, for example *zheng*, highlighted in (1b), and (iii) marked covertly (1c), i.e., taking the lack-viewpoint-morpheme (LVM) form (cf. Xiao 2002), as illustrated in the following examples from the CEPC parallel corpus (see unit 22.2 for a description of the corpora used in this work). Note that the Chinese examples are Romanized using Pinyin. They are followed by literal glosses and their English translations as they appear in the parallel corpus. The following abbreviations are used to gloss grammatical categories in Chinese examples: CLF (classifier), GEN (genitive), PSV (passive), PFV (perfective), PROG (progressive) and RVC (resultative verb complement).

- (1) (a) *zhe ben cezi liechu-le wuyong dupin de zhongzhong*  
this CLF booklet point-out-PFV misuse drug GEN various  
*qianzai wenti*  
potential problems  
'This booklet points out the potential problems from misusing drugs.'
- (b) *ruguo ni zheng bei chouqu xueye yangben...*  
if you PROG PSV take blood sample  
'If a blood sample is being taken...'
- (c) *zai chuxian aizibing zhengzhuang zhiqian, henduo ren dou bu zhidao*  
when appear AIDS symptom before many people all not know  
*ziji yi shoudao aizibing du de ganran*  
self already get AIDS virus GEN infection  
'Many people do not realise that they have been infected with HIV until they develop symptoms of AIDS.'

In (1a) perfectivity is marked by the aspect marker *-le*; in (1b) the adverb *zheng* indicates progressiveness. While Chinese is rich in aspect markers, it is interesting to note that covert marking of the LVM form is a frequent and important strategy used to express aspectual meanings in Chinese discourse, as shown in (1c). In this example, the verb phrase *shoudao...ganran* ‘be infected by’ shows a perfective meaning but is not marked overtly, even though the aspect marker *-le* can be used optionally in this context (i.e. *shoudao-le...ganran*). Such LVM forms typically occur in three situations. Stative situations normally take the LVM form because statives do not have to be marked aspectually (cf. Xiao 2002). For dynamic situations, there are two types of LVM sentences in discourse. They are imperfective as *irrealis* (e.g. future, habitual or conditional) or as having a perfective aspect marker deleted for discourse reasons (Chu 1987). In a discourse segment that presents a series of events, *-le* is ‘used as an explicit marker for the peak event’, i.e. the most important – and typically the last – event in a series (Chang 1986: 265), even though *-le* applies to the whole series of events rather than just the peak event as shown in (2a). Events other than the peak event can also be marked explicitly if the speaker chooses to regard them as separate events or wants any of them to stand out as separate events, as shown in (2b) (cf. Yang 1995: 138). For example (verbs in boldface, peak event underlined):

- (2) (a) *houlai, ta kaishi dui saiche fasheng xingqu, bingqie you-le*  
 then he start in banger-racing happen interest and have-PFV  
*yi-ge wengu de nüpengyou* (CEPC-health)  
 one-CLF steady GEN girlfriend  
 ‘Then, he started getting interested in banger-racing and got a steady girlfriend.’
- (b) *ta zhujian jianshao-le xiuxi rongji, tongshi ta de jiankang ye*  
 he gradually reduce-PFV sniff solvent meanwhile he GEN health also  
*you-le gaishan* (CEPC-health)  
 have-PFV improvement  
 ‘Gradually he sniffed less and less, and his health improved.’

In this case study we will first use ParaConc (version 1.0) to explore an English-Chinese parallel corpus and examine how aspectual meanings in English are translated into Chinese. We will then compare the translated Chinese texts with L1 Chinese texts to study the translation effect, followed by a contrast of texts from different domains and text types to see what effect, if any, domains and text types have on aspect marking. For a further discussion of this research question, readers can refer to McEnery and Xiao (2002), on which this case study is based.

While this study is focused on a non-European language, it should be apparent that the techniques and findings of this case study are also applicable to European languages. Since the corpus has been properly annotated, even if you do not understand Chinese, you will be able to understand the points raised in this case study and should be able to reduplicate the work presented in the following sections.

## 22.2 The corpus data

We will use three corpora for this work. The first corpus, the English-Chinese Parallel Health Corpus (CEPC-health), is aligned at the sentence level (see unit 5.3 for a discussion of alignment). It was constructed using a collection of English-Chinese bilingual pamphlets and leaflets issued from 1992 to 1994 by the Department of

Health and the Central Office of Information of the British government. The corpus covers one domain, public health, and two text types, exposition and narration. The English texts in the parallel corpus were annotated for parts-of-speech using the CLAWS tagger (see unit 4.4.1). The Chinese data was tokenized and POS tagged manually following a tagging scheme which was developed by Piao (2000) on the basis of the CKIP (Chinese Knowledge Information Processing) tagset (see Chen et al 1994).

As the frequency of aspect markers is exceptionally low in translated Chinese (cf. unit 22.3), we constructed a comparable L1 Chinese corpus, the Chinese Health Corpus (C-health), to compare the translated texts with L1 Chinese data (cf. unit 22.4). The corpus contains texts from current official websites for public health in China. As we hypothesize that the distribution of aspect markers may vary across domain and text type (cf. units 22.5 and 22.6), C-health matches the CEPC-health corpus in the domain and text types. To ensure the maximum comparability between the L1 Chinese corpus and the parallel corpus, this comparable corpus was POS tagged using the CKIP tagger. In addition to POS tagging, these two corpora have also been annotated semantically using a problem-oriented annotation scheme (see unit 22.3). They will also be used to study the effect of text types on aspect marking.

A third corpus, the Weekly corpus, will be used to explore the possible effects of domains and text types on the distribution of aspect markers in Chinese. This is an L1 Chinese corpus built with texts sampled from the *Southern Weekly*, a weekly newspaper with a sales volume of 1.3 million copies, published in China (see Xiao and McEnergy 2004b for a description). This newspaper corpus contains a mix of domains and text types. It was also tagged using the CKIP tagger. Table 22.1 compares the corpora used in this case study.

Table 22.1 Corpora used in this study

Corpus	Language	Domain	Tokens	Text type
CEPC-health	L1 English	Public health	31,638	2/3 exposition
	L2 Chinese		35,877	1/3 narration
C-health	L1 Chinese	Public health	34,174	2/3 exposition 1/3 narration
Weekly training	L1 Chinese	Mixed	96,897	Mixed
Weekly test	L1 Chinese	Mixed	10,054	Mixed

While the original versions of the Chinese corpora contain Chinese characters, we have converted these into the Roman alphabet using a system called Pinyin as we assume that most readers of this book will be unable to read Chinese characters. By using Pinyin, one can read Chinese without having to learn thousands of Chinese characters.

We assume that you have downloaded the three corpora used in this case study from our companion website and decompressed them into the following directories on your computer:

CEPC-health c:\My corpora\CEPC-health  
 C-health c:\My corpora\C-health  
 Weekly c:\My corpora\Weekly

### 22.3 Translation of aspect markers

In this section, we will examine how aspectual meanings in English are translated into Chinese. Unit 22.3.1 explores the translation of the progressive; units 22.3.2 and 22.3.3 are concerned with the perfect and the perfect progressive; unit 22.3.4 discusses the simple aspect.

### 22.3.1 The progressive

This section examines how the progressive in English is translated into Chinese. To find out the progressive forms in the English texts and their Chinese expressions in CEPC-health, do the following:

1. Activate ParaConc. You will see the interface of the concordancer, which is ready to load corpus files. The default number of parallel texts is two and English is selected by default. If you are running ParaConc for the first time and you have not already done so, a second language is undefined, as shown in Fig. 22.1.
2. Define the second language as Chinese (PRC) by clicking on the button next to *<undefined>* and selecting the language from the list, as shown in Fig. 22.2.
3. Click on the *Add* button on the left panel (for English). The *Select file(s) to open* window will appear (Fig. 22.3).
4. Locate the directory in which the CEPC-health corpus is stored and select the 20 files the filenames of which start with *hlthe* in the subdirectory named *English*. Press *Open* and the English texts of CEPC-health will be loaded into the concordancer. Do the same for the Chinese texts in the subdirectory named *Chinese* (Fig. 22.4).
5. Now ParaConc has loaded the English-Chinese parallel corpus and is ready for concordancing. Select *Search* → *Search* from the main menu. A new window will appear for you to type in a search pattern. Make sure that English is selected as the language option. Select *text search* for *Search Syntax*, and *Ignore case of letters* and *Sentence mode* for *General Search Control*, as shown in Fig. 22.5.
6. The default context is 40 characters on both sides of the search string. If you want to read more context, you can change this default setting by clicking on the *Option* button and defining *Context type* as *words* and the *size* as 10, as shown in Fig. 22.6.
7. In this case study, we are only interested in the progressive form in which the verb *BE* and present participles are separated by 0 – 2 words. First you need to press the *Option* button to define the special character @ as matching 1 and 2 words (see Fig. 22.6). Then type in *\*\_VB\* \*\_V?G* to get the concordance lines for the pattern *BE verb-ing*, and *\*\_VB\* @ \*\_V?G* to get the concordance lines where *BE* is separated by one or two words.

While in this step, we should have been able to define the special character @ as 0 – 2 words, which enables us to extract the structure of *BE* separated by 0 – 2 words from the present participle at one go, ParaConc version 1.0 does not work in this way. In the new version of the package, though, the special character @ can be defined as 0. Here the special characters \* and ? are used as wildcards (see case study 2). You will get 84 concordances for the first search string and 38 for the second search string. But these concordance lines need to be evaluated manually because there are eight instances of the perfect progressive, which will be discussed separately in unit 22.3.3. The concordances also include 30 invalid matches, as exemplified in (3).

Alternatively, you may wish to take advantage of the semantic annotation in the corpus, which we will discuss shortly, to avoid the time-consuming manual evaluation.

- (3) (a) [...] make sure you heat it until it **is piping hot**...
- (b) [...] the experience may **be damaging** to someone with mental illness...
- (c) A specific danger with mushrooms **is picking** the wrong ones...
- (d) Be fun, be consistent, **be caring**, but don't give the impression you approve of what they are doing.

In addition to its canonical use to signal the ongoing nature of a situation, the English progressive can refer to a habitual situation or an anticipated happening (cf. Leech 1971; see unit 10.6). The three types of progressives are annotated in the corpus respectively as <+PROG>, <+HABIT> and <+FUTURE>. As an alternative to extracting all progressive forms and evaluating them manually, readers can use this annotation scheme to extract English progressives of different types. To find the frequency and concordance lines of the progressive forms that denote on-going situations, simply type in *PROG* as the search string. There are 79 matches, as shown in Fig. 22.7. The upper window shows the matched concordance lines in the English texts while the lower window shows their Chinese translations. Type in the search strings *HABIT* and *FUTURE* respectively to get the frequencies and concordance lines of progressive forms that denote the habitual (four matches) and future meaning (one match).

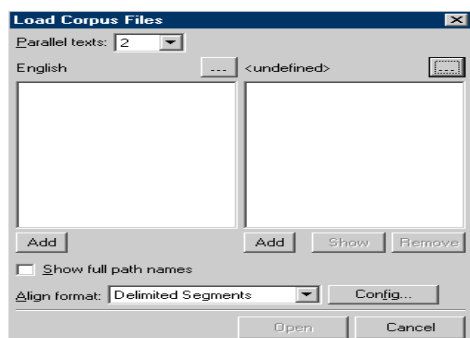


Fig. 22.1 The ParaConc interface

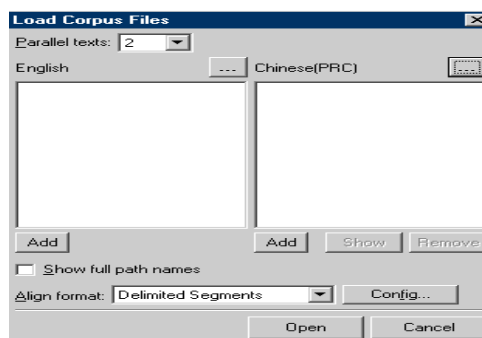


Fig. 22.2 Selecting the languages

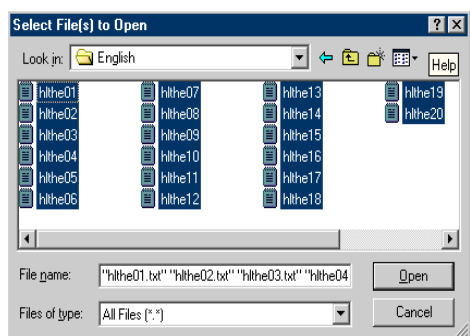


Fig. 22.3 Selecting files

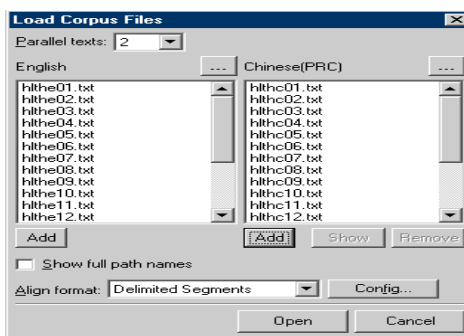


Fig. 22.4 Loading corpus files

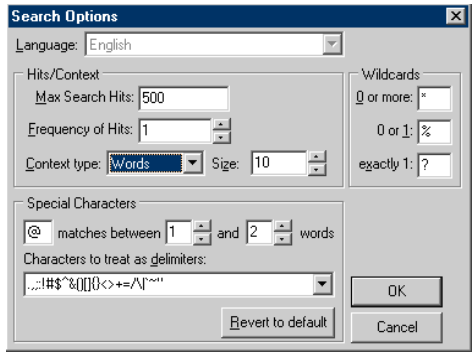
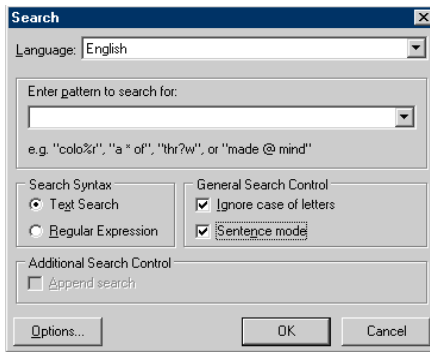


Fig. 22.5 The search window

Fig. 22.6 Defining search context

When you click on a concordance line in the upper window, the corresponding translation will be highlighted in the lower window. In this way, you can examine how a progressive form in English is translated into Chinese. To help readers who do not understand Chinese, we have annotated the Chinese texts in the parallel corpus with both part-of-speech and semantic information. The progressive meaning in Chinese texts is marked by the progressive *zai* (tagged as PROGZ), *zai* functioning both as a progressive marker and as a locative preposition (PROGZ2), the adverb *zheng* (ZHENG) and the durative *-zhe* (DURZ). The non-progressive meaning is marked by the actual *-le* (ACTL), the experiential *-guo* (EXPG), resultative verb complements (RVCC, RVCS, RVCD), or negation (NEG). In addition to the part-of-speech tags, there is an extra semantic tag in Chinese translations indicating whether an English progressive form is translated as progressive, non-progressive, or is paraphrased, i.e., an English progressive form is paraphrased with a noun phrase that is irrelevant to aspect marking. This semantic tag also shows how the aspectual meaning in the English source texts is marked in translations. For example, the tag <+PROG\_ZAI> means that a progressive form in English is translated as progressive and is marked the progressive marker *zai*. The tag <-PROG\_LVM> indicates that a progressive form in English is translated as non-progressive and is covertly marked, and <0PROG\_PARA> means that a progressive form in English is paraphrased in translation and is irrelevant to progressiveness.

8. Now you can examine the Chinese translations and classify each concordance line with the help of the tags described above. The results of your classification should match those in Table 22.2.

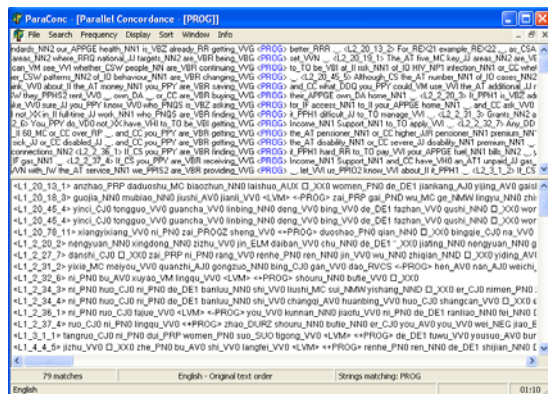


Fig. 22.7 Concordances of ongoing progressives

Table 22. 2 The translation of the progressive

Progressive translated as	Aspect marking in Chinese	Frequency	Marked /LVM
Progressive (56%)	Marked	(zheng)zai	14
		(zheng...)-zhe	3
		zheng	5
	LVM	25	0.88:1
Nonprogressive (35.7%)	Marked	-le	3
		RVC	4
		Negation	1
	LVM	22	0.36:1
Paraphrase (8.3%)		7	
Total (100%)		84	

Table 22.3 shows that if a progressive in the English source text refers to a habitual situation, the progressive aspect marker is not used in the Chinese translation, as shown in (4a). Similarly, when the progressive form in the English source text refers to an anticipated event, the progressive *zai* is not used in the Chinese translation, as shown in (4b). We hypothesize that this is because the progressive in Chinese only corresponds to the canonical use of the English progressive. While the progressive meaning is necessarily marked by the progressive form in English, it can be marked overtly or take the LVM form in Chinese. Our finding is in line with Comrie's (1976: 33) observation that 'in some languages the distinction between progressive and nonprogressive meaning by means of progressive and nonprogressive forms is obligatory whereas in others the use of the specifically progressive forms is optional.' It is clear that English belongs to the first type while Chinese belongs to the second type.

Table 22.3 Types of English progressives and their translations

Aspectual meaning in English	Translated in Chinese as	Frequency
Progressive	Progressive	47
	Non-progressive	25
	Paraphrase	7
Habitual	Non-progressive	4
Future	Non-progressive	1
Total		84

- (4) (a) If your child's health or behaviour shows that they **are taking drugs regularly**, you must take further action. (CEPC-health)  
*ruguo ni de haizi de jiankang huo xingwei xianshi tamen jingchang*  
 if you GEN child GEN health or behaviour show they regularly  
*shiyong dupin, na ni jiu bixu caiqu jinyibu de xingdong*  
 use drug then you just must take further GEN action
- (b) Graham (aged 12) seemed to be sniffing to relieve the tension he felt because his parents' marriage **was breaking up**. (CEPC-health)  
*Amote (shier sui) xiuxi rongji sihu shi weile jietuo youyu*  
 Graham (12 year) sniff solvent apparently is in-order-to relieve because-of  
*ta fumu hunyin polie er ganshou-dao de jinzhang*  
 he parent marriage break-up then feel-RVC GEN tension

### 22.3.2 The perfect

The English perfect relates a previous situation to the present. Note that as non-finite verb constructions with *having* plus a past participle and perfect forms taking modals like *must*, *would*, *could*, and *may/might* do not denote perfect meaning (cf. Comrie 1976: 55; Brinton 1988: 248), they are not discussed in this study. According to Comrie (1976: 56-62), the current relevance of a situation in the past can be manifested in four different ways: the perfect of result (5a), of experience (5b), of recent past (5c) and of persistent situation (5d).

- (5) (a) John has arrived. (Comrie 1976: 56)  
 (b) Bill has been to America. (*ibid*: 59)  
 (c) I have recently learned that the match is to be postponed. (*ibid*: 60)  
 (d) I've shopped there for years. (*ibid*: 60)

It is clear that the perfect is not simply concerned with perfectivity. While the first three types of perfect are perfective, the last is imperfective in nature (cf. Mourelatos 1981: 195). Unlike English, Chinese does not have a grammatical marker for the perfect. While the sentence-final *le* in Chinese also indicates current relevant state (c.f. Li and Thompson 1981), it is clearly different from the English perfect. First, the Chinese *le* is not restricted to the present. Rather it can indicate current relevance relative to a past, present, or future time reference. A further contrast between the English perfect and the Chinese *le* lies in the fact that, on the one hand, the perfect can carry the experiential meaning whereas *le* cannot, while on the other hand, *le* can refer to an imminent change of state whereas the perfect cannot (cf. Xiao 2002). When a Chinese sentence takes both the perfective *-le* and the sentence-final *le*, it is translatable by the English perfect of persistent situation (c.f. also Henne et al 1977: 113), because the sentence-final *le* in combination with the actual *-le* denotes a previous situation continuing into the present.

The four types of perfect in the English texts are tagged respectively as <PERFECT1>, <PERFECT2>, <PERFECT3> and <PERFECT4>. In the Chinese translations, there are also tags that show how a particular type of perfect is translated, e.g., <PERFECT1\_LE>, <PERFECT2\_GUO>, <PERFECT3\_RVC>, <PERFECT4\_LVM> and <PERFECT1\_PARA>, which respectively indicate that the Chinese translation is overtly marked by the actual *-le*, the experiential *-guo*, an RVC or is covertly marked as LVM or paraphrased as a noun phrase.

Table 22.4 Translation of the perfect

Type of perfect	Frequency	Translated as	Frequency	Marked/LVM
Result	71 (65.74%)	-le	16	0.78:1
		Negation	5	
		RVC	8	
		LVM	37	
		Paraphrase	5	
Experience	17 (15.74%)	-guo	7	0.7:1
		LVM	10	
Recency	6 (5.56%)	RVC	1	0.2:1
		LVM	5	
Persistency	14 (12.96%)	Negation	1	0.09:1
		LVM	11	
		Paraphrase	2	
Total	108 (100%)			



Table 22.5 Comparison of 4 types of perfect

Type of perfect	Translated as	Frequency	Marked	LVM	LL (1 d.f.)
Result	-le	16	36	47	7.61
	Negation	5			
	RVC	8			
	LVM	37			
Experience	-guo	7	2	16	
	LVM	10			
Recency	RVC	1			
	LVM	5			
Persistency	Negation	1			
	LVM	11			

If you have unloaded corpus files or exited ParaConc, you will need to repeat steps 1 and 3 – 6 in unit 22.3.1, though there is no need to define languages in step 2 again. Otherwise, simply select *Search* → *Search* from the menu and type in *PERFECT1*. There are 71 matches. Now you can examine how the perfect of result is translated into Chinese by looking in the corresponding Chinese translations for the tag listed above. You will find that 16 translations take *-le*, five take negative adverbs, eight take RVCs, 37 take the LVM tag, and five are paraphrased. Do the same with the search strings *PERFECT2*, *PERFECT3* and *PERFECT4*. Your results will match those in Table 22.4. The table shows the frequencies of the four types of perfect in the English texts and their translations. It can be seen that the perfect of result is the most common of the 4 types of perfect. When perfect constructions are translated into Chinese, they tend to depend on context to indicate perfect meanings rather than mark it explicitly, though aspect markers such as *-le*, *-guo*, RVCs and the sentence-final *le* could be used where appropriate. Whether the translations take aspect markers or contextually imply the perfect depends largely on the type of perfect in the English source texts. While the LVM forms are dominant in all of the four categories, aspect markers are more frequent in translations of the perfect of result and the perfect of experience than the latter two categories. This difference is statistically significant at  $p < 0.01$  (the critical value for which is 6.64 with 1 d.f.), as shown in Table 22.5.

### 22.3.3 The perfect progressive

The perfect progressive is an interaction between the perfect and the progressive, indicating the immediacy of an effect. It takes the form of *HAVE been verb-ing*. Therefore, it is easy to find this structure in the parallel corpus. Simply type in the search string *\*\_VH\* been\_VBN \*\_V?G*, where the special character ? matches any single character so that *VVN*, *VDN*, *VHN* and *VBN* are included. In Chinese translations, there are also semantic tags that indicate whether the perfect progressive is translated as progressive or non-progressive. These tags also show how the aspectual meaning is marked in translated texts. For example, the tag <-PROG\_LE> means that the perfect progressive is translated as non-progressive and is marked by the aspect marker *-le*, and <?PROG\_LVM> indicates that the translation is covertly marked as LVM and is ambiguous between a progressive and non-progressive reading. There are only eight instances of the perfect progressive in the English texts of the CEPC-health corpus. Of these seven are translated as non-progressive and one instance (i.e. 'If you find that your child has been using drugs what can you do' translated as *ruguo ni fajue ni de haizi shiyong dupin ni ruhe yingfu*) is ambiguous

between a progressive and a non-progressive reading in the Chinese translation. Surprisingly, while Chinese has a well-established progressive marker *zai*, translations of the English perfect progressive normally preserve its perfect meaning and discard its progressive meaning.

This observation is contrary to what Zhang (1995: 181) would have us expect. Zhang uses the example in (6a) to assert that translations of the perfect progressive only render the progressive part of meaning. But the translation in (6b) is not as accurate as (6c) which does not support Zhang's claim. In this alternative translation, the addition of the temporal adverb *gang* 'just a while ago, just now' renders the immediacy effect expressed by the perfect progressive in the source text (i.e., the paint is still wet). Zhang's assertion, however, is clearly ungrounded, as the corpus example in (7) illustrates:

- (6) (a) They have been painting on my staircase. (Zhang 1995: 181)  
 (b) *wo jia de louti zhengzai youqi* (Zhang 1995:181)  
 my home GEN staircase PROG paint  
 (c) *wo jia de louti gang youqi-guo* (Our translation)  
 my home GEN staircase just paint-PFV
- (7) In most cases, your complaint can probably be handled by the member of staff you've been dealing with. (CEPC-health)  
*zai daduoshu de shijian zhong, ni de tousu keneng hui you*  
 in most GEN time during you GEN complaint probably can by  
*yi wei cengjing he ni jiechu de zhiyuan lai chuli*  
 one CLF once with you contact GEN staff-member come deal-with

Like the perfect and the progressive, translations of the perfect progressive primarily take the LVM form. In the data, only two instances are explicitly marked by *-le* (tagged as <LE>) while aspectual meanings are marked covertly (tagged as <LVM>) in all of the other translations, registering a marked/LVM ratio of 0.33:1.

### 22.3.4 The simple aspect

According to Hatcher (1951: 259-260), the simple form in English has no aspectual meaning and it is indifferent to aspect. As such, the simple aspect may express a perfective, habitual or timeless situation. When aspect interacts with tense, we have the simple past, the simple present and the simple future. The simple present prototypically refers to states, i.e., statements made for all time (8a) and habitual situations (8b).

- (8) (a) Flu is more common in the winter months. (CEPC-health)  
 (b) Septicaemia often occurs with meningococcal meningitis. (CEPC-health)

As stative situations do not have to be marked aspectually in Chinese to have a closed reading, translations of the simple present tend to take the LVM form. The simple future refers to a future time reference. Chinese does not mark tense grammatically (cf. Wang 1943: 151; Norman 1988: 163). Future time references in Chinese are most frequently expressed by modal auxiliaries (cf. Xiao and McEnery 2004b). As a predicate can only be marked for either mode or tense and aspect (Biber et al 1999), Chinese translations of the simple future do not take aspect markers. As the canonical use of the simple past is to locate a situation prior to the present moment, it is natural that perfectivity in English is most commonly expressed by the simple past, less often

by the simple present (Brinton 1988: 52). Therefore, in this case study, we will investigate how the simple past and the simple present are translated into Chinese. In this section, we will only include lexical verbs (tagged by CLAWS as *VVD* for the simple past and *VV0* and *VVZ* for the simple present in the English corpus) while excluding the verbs *BE* (*was/were/is/am/are*), *HAVE* (*had/have/has*) and *DO* (*did/do/does*), because these verbs can also function as auxiliaries and are likely to make the concordancing procedure too complicated for the purpose of this study. A further reason for this decision is that the verbs *BE* and *HAVE* are stative verbs when they have lexical meanings (see discussion in unit 22.4).

### 22.3.4.1 The simple past

To get the frequencies and concordances of the simple past form of lexical verbs, type in the search pattern *\*\_VVD*. There are 84 matches. You can examine the translation pattern of the simple past by looking at the semantic tags in angled brackets in the Chinese translations. For example, *<PAST\_+PERFECTIVE\_GUO>* indicates that a simple past form in English is translated as perfective and is marked by the aspect marker *-guo*. *<PAST\_-PERFECTIVE\_ZAI>* means that a simple past form is translated as imperfective and is marked by the aspect marker *zai*, while *<PAST\_OPERFECTIVE\_PARA>* shows that the Chinese translation is a paraphrase to which an aspectual analysis does not apply. By examining these tags in the Chinese translation of a corresponding concordance, you will be able to establish a translation pattern of the simple past. Here you will notice that verbs taking the simple past form are found primarily in texts Nos. 15 and 18, as shown in Fig. 22.8. For the moment, we will simply note this phenomenon, though we will return to consider this observation in unit 22.6.

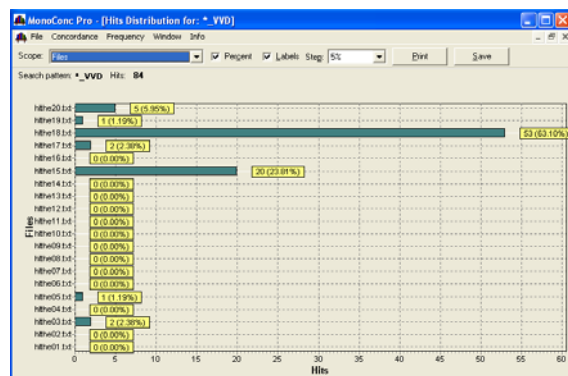


Fig. 22.8 The distribution of the simple past in CEPC

Table 22.6 Translation pattern of the simple past

Simple past translated as	Aspect marker in Chinese	Frequency
Perfective	-le	10
	Negation	1
	RVC	4
	-le + RVC	2
	LVM	63
Imperfective	-xiaqu	1
Paraphrase		3
Total		84

Table 22.7 Translation pattern of the simple present

Simple present translated as	Translated into Chinese as		Frequency
Perfective	Marked	-le	9
		-guo	1
		RVC	61
		Negation	2
		-le + RVC	3
	Negation + RVC	1	
	LVM		810
Imperfective	Marked	(zheng...)-zhe	10
		(zheng)zai	7
		zheng	1
		-qilai	1
Others	Paraphrase		35
	Omission		5
	As original		1
Total			947

Table 22.6 shows the translation pattern of the simple past. It is clear that verbs taking the simple past form are translated primarily as perfective (95%), either marked overtly or taking the LVM form. This is as expected, because the English simple past normally marks situations that were completed or terminated in the past. In Chinese, situations of this type are most frequently expressed by the actual aspect (either marked by *-le* or taking the LVM form) and the completive aspect (marked by RVCs), because the former indicates the actualization while the latter indicates the completiveness of a situation. It is also interesting to note that aspectual meanings are frequently marked covertly in Chinese discourse, even though Chinese is rich in aspect markers. If we discount the three instances of paraphrase, the translated texts register a marked/LVM ratio of 0.222:1.

#### 22.3.4.2 *The simple present*

As the simple present form of a lexical verb is tagged as *VV0* or *VVZ*, you will need to search for the corpus twice using the search patterns *\*\_VV0* and *\*\_VVZ*. There are 735 matches for the first search pattern and 212 matches for the second. In the Chinese translations, the annotation scheme for the simple present is the same as that used for the simple past with the exception that *PAST* is replaced by *PRESENT*. The annotation indicates whether the simple present is translated as perfective (+PERFECTIVE), imperfective (-PERFECTIVE), or is paraphrased or omitted in translations and is thus irrelevant to perfectivity (0PERFECTIVE). It also shows how the aspectual meaning denoted by the simple present is marked in Chinese translations, e.g., taking *-le*, *-guo*, RVC, or paraphrased or omitted. With the help of this annotation scheme, you will be able to see the general pattern of how the English simple present is translated into Chinese, as shown in Table 22.7. It can be seen from the table that, as with the simple past, when situations referred to by the English simple present are overtly marked in Chinese translations, they most frequently take RVCs and the actual *-le*. Since states do not have to be marked aspectually, the majority of situations expressed by the simple present take the LVM form in Chinese translations. If we discount the frequencies in the *Others* category, a much lower marked/LVM ratio (0.106:1) than that for the simple past is found in Chinese translations.

## 22.4 Translation and aspect marking

The discussion in unit 22.3 showed that the marked/LVM ratios in Chinese translations for English aspect markers are exceptionally low. We hypothesize that the relatively low frequency of aspect markers in translated texts is a result of translation choices and is atypical in L1 Chinese. In this section, we will test this hypothesis using a comparable corpus of L1 Chinese, C-health. As the distribution of aspect markers may vary across domain and text type (cf. units 22.5 and 22.6), C-health was constructed using roughly the same sampling frame as that used for the CEPC-health corpus. Like the CEPC parallel corpus, the comparable L1 Chinese corpus is also semantically annotated in addition to POS tagging.

To test the above hypothesis, we will experiment with two well-established perfective aspect markers in Chinese, *-le* and *-guo*, which are tagged in the C-health corpus respectively as *ACTL* and *EXPG*. Two additional tags were added to simplify the concordancing process. Verbs marked by *-le* or *-guo* were labelled as <MARKED> while the others are labelled as <LVM> unless they take a modal auxiliary.

It is a straightforward process to get the frequencies of marked cases in C-health using MonoConc Pro. Simply load the two corpus files named *narrative.txt* and *expository.txt* in the directory for the C-health corpus and type in the search string *MARKED* (see case study 2 for the use of MonoConc Pro). There are 213 matches.

The frequencies of marked cases (i.e., verbs taking with *-le* or *-guo*) in the translated texts in the CEPC parallel corpus can be obtained by loading the 20 corpus files whose filenames start with *hlthc* in the subdirectory *Chinese* (but remember to unload the C-health corpus first) and use the batch search function of MonoConc Pro, typing in the search patterns *\*\_ACTL* and *\*\_EXPG*. There are 98 concordances for the search strings.

Table 22.8 shows the frequencies of the two perfective markers in both corpora. As can be seen from the table, the normalized frequency of perfective markers is twice as high in the L1 Chinese comparable corpus as in the translated texts. The calculated log-likelihood (LL) score 49.113 is much greater than 10.83, the critical value for significance at  $p < 0.001$  (1 d.f.). In other words, the difference in the distribution of perfective aspect markers in L1 Chinese texts and in the translated texts is statistically significant.

A closer examination of the marked/LVM ratio in the Chinese translations of the tensed verbs in CEPC-health English texts illustrates this point well. Table 22.9 shows the distribution of tensed verbs in CEPC-health English texts. In the table the column *Marked* only includes the counts of *-le* and *-guo*, and the column *Others* includes the instances of paraphrase, omission and imperfective viewpoints in the Chinese translations. If we discount the instances in the *Others* column, the marked cases account for only 4.79% of the total instances of perfectives, registering a marked/LVM ratio of 0.05:1. This is in sharp contrast with L1 Chinese data in the C-health corpus, as shown in Table 22.10.

To match the counting procedure applied to CEPC-health English texts, verbs that take a modal auxiliary or do not function as predicates are not annotated and thus not counted. It can be seen from the table that the marked/LVM ratio in L1 Chinese texts is nearly twice as high as that in the translated texts. This difference is statistically significant, as indicated by the calculated log-likelihood score 15.792, which is considerably greater than the critical value 10.83 for significance at  $p < 0.001$ .

Table 22.8 Perfective markers in CEPC-health Chinese texts and C-health

Corpus	Tokens	Perfective markers	Frequency per 10K tokens	LL (1 d.f.)
C-health	34,174	213	62.33	49.113
CEPC-health	35,877	98	27.32	

Table 22.9 Translations of tensed verbs in CEPC-health

Aspect	Marked	LVM	Others	Total
Simple present	13	874	60	947
Simple past	12	68	4	84
Perfect	23	78	7	108
Progressive	3	27	54	84
Perf. Progressive	2	6	0	8
Total	53	1,053	125	1,231

Table 22.10 Predicates in L1 Chinese texts

Corpus	Marked	LVM	Marked/LVM	Marked %	LL value
CEPC-health	53	1,053	0.05:1	4.79%	15.792
C-health	213	2,329	0.09:1	8.38%	

Note that the frequencies obtained from C-health include all predicates, whether they are stative or dynamic. The frequencies from the CEPC-health corpus, however, include only lexical verbs. If we include verbs *BE* and *HAVE*, a more marked contrast between translated texts and L1 Chinese texts is expected, because the two verbs are most frequently translated as statives and take the LVM form in translations. Furthermore, verbs taking the LVM form in translations for the progressive fall under the category of *Others* in Tables 22.9 and 22.10, whereas the frequency of LVM cases from C-health includes all verbs taking the LVM form, irrespective of whether an LVM form denotes a perfective or imperfective meaning, because such information is not annotated in the corpus. If we classified LVM verbs that are used imperfectively as *Others*, the contrast between translations and L1 texts would have been even more marked.

## 22.5 Domain and aspect marking

Table 22.10 shows that the frequencies of perfective aspect markers in both C-health and the translated texts in CEPC-health are rather low. Even if we discount the effect of translation on the distribution of aspect markers and consider L1 Chinese data alone, there are only 62.33 occurrences of *-le* and *-guo* per 10,000 tokens (see Table 22.11). As Chinese is an aspect language and hence aspect markers are supposed to occur more frequently, we hypothesize that the distribution of aspect markers varies across domain.

Table 22.11 Frequencies of perfective aspect markers

Corpus	Tokens	Raw frequency	Frequency per 10K tokens
Weekly training	96,897	1,117	115.28
Weekly test	10,054	132	123.29
C-health	34,174	213	62.33

Table 22.12 The Weekly training vs. the Weekly test corpus

			Size_Marker		Total
			Fre. of markers	Corpus size	
Corpus	Test	Count	132	10054	10186
		Expected Count	117.6	10068.4	10186.0
	Training	Count	1117	96897	98014
		Expected Count	1131.4	96882.6	98014.0
Total		Count	1249	106951	108200
		Expected Count	1249.0	106951.0	108200.0
Pearson chi-square	1.975		Log-likelihood (LL)		1.907

Table 22.13 The Weekly corpus vs. the C-health corpus

			Size_Marker		Total
			Freq. Of markers	Corpus size	
Corpus	C-health	Count	213	34174	34387
		Expected Count	352.6	34034.4	34387.0
	Weekly	Count	1249	106951	108200
		Expected Count	1109.4	107090.6	108200.0
Total		Count	1462	141125	142587
		Expected Count	1462.0	141125.0	142587.0
Pearson chi-square	73.576		Log-likelihood (LL)		82.087

In this section, we will test this hypothesis by comparing the distribution of *-le* and *-guo* in the C-health corpus and the Weekly corpus. C-health is confined to only one domain, i.e. public health, while the Weekly corpus covers mixed domains. We already have the frequency of the marked cases in the C-health corpus. To get the frequencies of *-le* and *-guo* in the Weekly training and test corpora, simply load the two corpora separately (but again, remember to unload the previous texts first) and type in *\*\_ACTL*, *\*\_DUAL*, and *\*\_EXPG*, where *DUAL* is the POS tag for *-le* functioning as both an perfective marker and a change-of-state (COS) or sentence-final *le*. There are 1,019 matches of *ACTL*, 23 matches of *DUAL* and 75 matches of *EXPG* in the training corpus, and 119 instances of *ACTL*, four instances of *DUAL* and nine instances of *EXPG* in the test corpus. Table 22.11 gives the frequencies of *-le* and *-guo* in the corpora.

As Tables 22.12 and 22.13 show, while the two perfective markers are slightly more frequent in the test corpus than in the training corpus, the difference is not statistically significant (the LL value is 1.907, less than 3.84, the critical value for significance at  $p < 0.05$ ). In contrast, the difference between Weekly as a whole and C-health is statistically significant (the LL value is 82.087) at  $p < 0.001$ , the critical value for which is 10.83. Perfective aspect markers appear in the Weekly corpus nearly three times as frequently as in the C-health corpus.

## 22.6 Text type and aspect marking

The perfective aspect is expressed most commonly by the simple past in English (cf. Brinton 1988: 52). As you may have noticed in unit 22.3.4.1, the majority of the verbs taking the simple past form were found in only two texts: *Drugs – a Parent’s Guide (hlthe15.txt)* and *Solvents – a Parent’s Guide (hlthe18.txt)*. These two texts constitute around one third of the English data in the parallel corpus and are primarily narrative in nature, showing what certain parents did to help their children stop their drug and solvent abuse. The other two thirds of the English texts in the corpus are expository

writings and only contain a dozen instances of *VVD*. As such, we hypothesize that different text types may have an influence on aspect marking. In this section, we will test this hypothesis by examining the relation between the frequency of simple past forms and the number of tokens in the two text types in the English data and the relation between verbs taking *-le* or *-guo* in the two text types in the C-health corpus. To discover the frequency of verbs taking the simple past form in the narratives in the parallel corpus, you need to initialize the concordancer by unloading previous corpus files, and then load files named *hlthe15.txt* and *hlthe18.txt* in the subdirectory *English* in the file folder containing the CEPC-health parallel corpus. Type in the search pattern *\*\_VVD*. There are 73 matches. Do the same with the other files in that subdirectory. You will find only 11 matches. This means that the narrative texts that make up one third of the English texts in the parallel corpus account for 86.9% of the total frequency of *VVD* while the expository texts that make up two thirds of the English data account for only 13.1%.

It can be seen from Table 22.14 that the simple past occurs more than 12 times as frequently in narrative texts as in expository texts. The calculated log-likelihood score 65.216 is considerably greater than the critical value of 10.83 for significance at the level  $p < 0.001$ , indicating that the two text types are indeed different (cf. unit 15.3). As perfectives in English are all marked, the marked/LVM distinction is irrelevant. In Chinese, however, perfectives can be overtly marked or take the LVM form in discourse. To find out if the above finding also applies to the Chinese data, we will first search the two components of the C-health corpus to get the frequencies of marked cases and LVM cases. Load the corpus file named *narrative.txt* in the directory for C-health and type in *MARKED* and *LVM* respectively. There are 94 instances of marked cases and 526 instances of LVM. Do the same for *expository.txt* in the same directory. There are 119 marked cases and 1,803 instances of LVM.

Table 22.14 Text type vs. aspect marking in English (simple past)

Text type	Tokens	Raw frequency	Frequency per 10K tokens	LL value (1 d.f.)
Narrative texts	11,226	73	65.03	65.216
Expository texts	20,512	11	5.36	

Table 22.15 Text type vs. aspect marking in Chinese (*-le* and *-guo*)

Text type	Narrative texts	Expository texts
Number of tokens	7,167	27,007
Number of predicates	620	1,922
Raw frequency of markers	94	119
Frequency per 10K tokens	131.16	44.06
Marked percent	15.16%	6.19%
Marked/LVM ratio	0.179:1	0.066:1

Table 22.16 Text type vs. aspect marking (LL test)

Text type	Marked	LVM	LL ratio (1 d.f.)	Sig. level
Narrative	94	526	43.656	$p < 0.001$
Expository	119	1,803		

Table 22.15 shows the frequency of perfective aspect markers *-le* and *-guo* in the C-health corpus and other relevant statistics. In spite of the overall low frequencies of



perfective markers in the domain of public health, there is a marked contrast between the two text types in the domain. Perfective markers occur 2.97 times as frequently in the narrative texts as in the expository texts (per 10,000 tokens). Both the proportion of marked cases and the marked/LVM ratio are more than twice as high in the narrative texts as in the expository texts. The LL test in Table 22.16 shows that the difference in the distribution of aspect markers in the two text types is indeed statistically significant.

In this section, we only experimented with two text types. However, our finding, that the distribution of aspect markers varies across text type, appears to be generalizable to other text types. Li, Thompson and Thompson (1982: 26-27), for example, also observe that the 'perfect' particle *le* is 'very rare in expository and scientific writing and practically non-existent in news-reporting, speeches, lectures, and proclamations.' Hence, we can claim that text types do indeed have an effect on aspect marking (see unit 15.3).

## 22.7 Unit summary and suggestions for further study

In this unit, we investigated how aspectual meanings in English are translated into Chinese and explored the effects of translation, domains, and text types on aspect marking. We used one English-Chinese parallel corpus and two L1 Chinese corpora (one comparable corpus and one reference corpus) in this unit to produce a number of findings based upon this data.

It was found that the English progressive can be translated into Chinese either as ongoing or as perfective. If a progressive form in an English source text refers to a habitual situation or an anticipated happening, the progressive aspect marker is not used in Chinese translation. We hypothesize that this is because the progressive in Chinese only corresponds to the canonical use of the English progressive.

Also, in English, the perfect of result is the most common of the four types of perfect identified by Comrie (1976). When perfect constructions are translated into Chinese, they more often than not depend on context to indicate the perfect meaning. This is because Chinese does not have a grammatical aspect marker for the perfect. In this case, however, aspect markers such as *-le* and *-guo* could be used to mark the perfective. Whether the translations take overt aspect markers or imply the perfect meaning contextually depends largely on the type of perfect in the English source text. With reference to the simple aspect, we found that situations marked by the English simple aspect are mainly presented perfectively and most of them take the LVM form in Chinese translations. The high frequency of perfectives in translations of the simple aspect can be accounted for by the fact that the simple forms in English, the simple past in particular, are basically perfective in nature (cf. Brinton 1988). Translations of the simple past show a marked/LVM ratio twice as high as that in translations of the simple present. A natural explanation for this contrast is that the simple present typically denotes states, which do not have to be marked aspectually.

The use of translated data allowed us to discover that translated Chinese texts differ markedly from L1 Chinese texts with regard to aspect marking. The frequency of the perfective markers *-le* and *-guo* is nearly twice as high in L1 Chinese data as in translated Chinese texts. We hypothesize that this difference is a consequence of the potential influence of the translation process, since translators cannot rid themselves of the influence of the source language when translating.

Considering the L1 Chinese health corpus data in contrast to the other forms of data, it is apparent that in Chinese the frequency of aspect markers in the domain of public

health is considerably lower than in the mixed domains. The perfective aspect markers *-le* and *-guo* distribute nearly three times as frequently in general domain as in the domain of public health. This can be explained by the fact that the domain of public health primarily contains texts of an expository nature and is concerned with general truth.

Finally, the distribution of aspect markers varies significantly between different text types even in the same domain. Perfective markers are more frequent in narratives than in expository writings. This is because narrative texts typically present situations that are actualized, experienced or completed in relation to a particular reference time whereas expository writings are primarily concerned with timeless states.

The above findings demonstrate the potential values of aligned parallel corpora and comparable L1 corpora in translation and contrastive language studies.

Methodologically, this unit also showed you how to explore parallel corpora using ParaConc and provided you with an opportunity to practise using MonoConc Pro with monolingual corpus data. At this point you might wish to explore the CEPC-health parallel corpus to find out how the English simple future is translated into Chinese.

The simple future in English is marked by *will*, *shall*, *be going to* and *be (about) to*. Will your observations support our findings concerning the effect of translation, domain and text type on aspect marking?

