A Parallel Corpus-based Study of Translational Chinese

WANG Keifei
Beijing Foreign Studies University

QIN Hongwu
Qufu Normal University

Abstract: This paper, based on a Chinese-English parallel corpus, probes into the features of translational Chinese, and has made the following findings: 1) Different from what is generally believed, translational Chinese, compared with original Chinese, has higher type-token ratio and longer sentence segments; 2) there is difference between original Chinese and translational Chinese in POS distribution; the former uses more function words and fewer content words; 3) translational Chinese tends to exaggerate the compositional potentiality of some words or morphemes, which results in high-frequency use of some lexical bundles. The above features specific to translational Chinese cannot find full expression in universals of translation. Considering this, we need to look to interference from source language (English) for an adequate explanation.

Key Words: parallel corpus; translational Chinese; contrastive analysis

1. Introduction

Translational language (TL) retains, to varying degrees, some features of its source language (SL), and it is "a non-standard version of the target language that is […] affected by the source language" (Hopkinson 2007). By 'non-standard', we mean the language used in translation is not as idiomatic and prototypical as it is in original texts in the same language, because the former contains deviations from the typical patterns of TL, with SL being its origin (Toury 1995: 208).

Features of translational language can be observed in many ways. According to Santos (1995: 60), we can go into 1) properties of all translation, i.e. the universals of translation (Baker 1993), 2) properties of translations particular to a source language and target language, i.e. translationese, and 3) properties of particular translated texts (with the author and the translator taken into consideration). This study focuses on the second set of properties.

The use of parallel corpus in the study of translational Chinese has emerged in recent years in China, as has been done by Ke Fei (2003), Qin & Wang (2004). However, studies in this field are rare and inconsistent, owing to the lack of reasonable methodology and appropriate tools. For a better inquiry into translational Chinese, this study attempts a multilevel analysis.

2. Universals of Translation

Baker (1993, 1998a) reports that all translational languages share some features, namely, i) simplification (including lexical simplification, syntactic simplification, stylistic simplification); ii) explicitation (what is implied in original text gets explicit, so are the cohesion markers); and iii) normalization (source-text textemes tend to be converted into target-language repertoremes and diversity is lost).

Baker (1998b:225) does acknowledge that in translation, some stylistic features of the source text tend to be transferred to the target text; however, her findings are made mainly on the basis of monolingual comparable corpus, without duly taking into account the influence of source language (see Hansen & Teich 2001; Wu & Huang 2006; Huang & Wang 2006). Moreover, Baker's study revolves around shallow linguistic features such as word length, type/token ratio, sentence length, lexical density, etc.; as a result, many abstract features peculiar to the languages in translation pair have been blinked.

Considering the empirical data on which the universals are based, it is questionable whether

they are applicable to all translational languages, especially to languages of different types. For instance, we are quite concerned whether they are applicable to translational Chinese. To observe their applicability, we employ a multilevel analysis approach to translational Chinese. On the one hand, we make use of Baker's analytic techniques; on the other hand, we conduct micro-level analysis (like POS distribution, compositionality and load capacity) so as to arrive at an adequate description.

3. Normality, corpus and multilevel analysis

Normality:

Chinese has its own typical linguistic patterns, and this typicality is termed 'normality' (Yu 2002:151). Normality is not a set of rules, but a language intuition which cannot be precisely measured or defined. However, we can assume its presence in a certain amount of original Chinese texts, because, compared with translated Chinese texts, the former is closer to the normality.

Corpus

The comparable corpus we use for description and analysis are from the General Chinese-English Parallel Corpus (GCEPC)[1] created by Beijing Foreign Studies University(BFSU). GCPEC has four subcorpora, namely Chinese-English Literature, Chinese-English Non-literature, English-Chinese Literature, English-Chinese Non-literature. The Chinese texts taken from CE and EC corpora can form comparable corpus, and the same is true for English texts. Besides, GCEPC enables us to take into account the English source texts in analysing translational

[1] GCEPC is a Chinese-English bidirectional parallel corpus. The corpus-building project was led by Professor Wang Kefei and was completed in 2004. It contains about 20 million words and characters which are stored in XML format. The corpus for this study is about 3.5 million words and characters.

Chinese.

Multilevel analysis

This study first analyses translational Chinese at a macro level (TTR, word length, sentence length), and then turns to micro-level analysis (POS distribution, keywords analysis, compositionality and lexical bundles).

4. Macro-level Description

The Macro-level description mainly concerns type-token ratio (TTR), word length count and sentence length count that reflect the difference between translated and original Chinese texts. Chinese words for TTR and sentence length calculations are segmented by ICTCLAS. Below is the data extracted from Wordlist in WordSmith 4.0.

Table 1 Macro-level description of translated and non-translated Chinese texts

	TOKENS (Counted)	TYPES	STTR	Word Length	S Length	SS Length
OCT (lit)	466,414	23,047	46.72	1.36	25.46	6.02
OCT (non-lit)	222,758	11,066	41.92	1.76	27.05	7.20
OCT	**689,172**	**28,437**	**45.19**	**1.49**	**25.95**	**6.35**
TCT (lit)	578,148	24,213	47.36	1.44	25.81	7.00
TCT(non-lit)	496,218	26,174	47.65	1.64	31.52	8.58
TCT	**1,074,366**	**36,354**	**47.49**	**1.53**	**28.27**	**7.65**
EST(lit)	546,632	22,409	43.21	4.26	16.76	6.79
EST (non-lit)	487,673	25,739	44.37	4.87	20.24	9.32
EST	**1,034,305**	**35,695**	**43.75**	**4.54**	**18.23**	**7.78**

As shown in Figure 1, OCT(original Chinese texts), TCT(translated Chinese texts) and EST(English source texts) are different in type-token ration, word length, and sentence segment length (*SS Length* for short). What follows are details.

STTR

Generally, the larger the corpus is, the smaller the TTR. Given the obvious difference of the subcorpora in size, we cannot stop at a simplified TTR count, what we really need is STTR (standardized type-token ratio)[1] count.

Higher STTR indicates more different lexical items, while lower STTR suggests that fewer specific words are used and the more general ones are frequent (Westin 2002:75). As is shown in Table 1, STTR for OCT is 2.3 percent lower than that for TCT (47.49 vs 45.19), however, they are higher than that for EST (43.75). This difference suggests TCT, in comparison with OCT, is not so 'simplified' in terms of lexical diversity. This may serve as counter evidence of lexical simplification which states translational language tends to use simple words for ease of understanding.

Word Length

For English and for many other alphabetic languages, word length is a way of measuring lexical specificity and diversity. For Chinese, however, word length count can reflect idiomaticity of language use: in Mandarin Chinese, most words used in Chinese discourse are disyllabic and monosyllabic, but "monosyllabic words are most frequently used" (Lü 1981:9).

The mean word length of TCT and OCT are similar, only that the former is 0.04 longer than the latter. Moreover, in contrast with TCT (53.16% words being monosyllabic, and 41.72% words disyllabic), OCT uses more monosyllabic words (56.96%), and fewer disyllabic words (38.58%). This suggests that TCT is not as idiomatic as OCT in word length.

Sentence Length (S Length) and Sentence Segment Length (SS Length)

Sentence length measures sentences that begin with capital letters (for English but not for

[1] i.e. the ratio is calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of the subcorpus in question.

Chinese) and ends in full stops, exclamation marks, question marks or colons. We hereby note that Chinese sentences are calculated in words, but not in characters.

The calculation of mean S length yields the following results: TCT uses longer sentences (2.32 more words on average) than OCT; in addition, TCT uses much longer sentences than EST (18.23 for EST, while 25.81 and 28.27 respectively for OCT and TCT). The following instances can show why Chinese uses longer sentences than English does.

（1）And Pinkerton--Pinkerton--he has collected ten cents that he thought he was going to lose .

那么/c 平克顿/nr --/x 平克顿/nr --/x 他/r 一定/d 是/u 要/v 回来/dg 一/m 角/q 钱/n 的/u 老/a 账/n ，/w 这笔/r 钱/n 他/r 本来/d 以为/v 没有/v 盼头/n 了/y 。/w （16：22）

then_c Pinkedun_nr --_x Pinkedun_nr --_x he_r surely_d is_u claim_v back_dg one_m cent_q money_n DE(gen)_u old_a debt_n， _w this_r sum_c money_n he_r originally_d believe_v not-have_v prospect_n LE(crs)_y 。_w

（2）You speak collectedly, and you--are collected.

你/r 这/r 话/n 倒/y 还有/v 自制力/n ，/w 而/c 你/r --/x 也/y 确实/d 镇静/a 。/w （7：10）

your_r this_r utterance_n nevertheless_y have_v forbearance_n， _w but_c you_r --_x also_y really_d calm_a 。_w

Chinese, an isolating language, usually resorts to lexical means to express what is expressed grammatically in English. For example, the relative pronoun *that* in sentence (1) is replaced by a noun phrase 这笔钱 (literally *this sum money*), and the added expression本来 (literally

originally) serves to express the temporal meaning in *thought* (past tense).

The above amplification is to some degree compulsory, while a lot other amplifications are optional. For instance, as shown in example 2, there is shift of part of speech in translation [e.g. form *speak* to 这话(lit. *this utterance*); from *collectively* to 有　自制力(lit. *have forbearanc*e)], and this shift forces the other sentential elements to change accordingly. More than that, modality implied in the original is made explicit in translation [such as 倒 (*nevertheless*); 也 (*also*); 确实(*really*)]. Of whatever type, amplification contributes to the expansion of sentence length, and it is just this amplification that makes sentences in TCT much longer than their counterparts in EST, a support for explicitation in translation.

A simple S length calculation could not reveal intro-sentence properties of a language, so the result it derives might prove to be mistaken if we take sentence segment into account.

As Chen (1994) reports, "about 75% of Chinese sentences are composed of more than two sentence segments[1] separated by commas or semicolons." If this is true, SS length calculation might tell us more about the organization of Chinese sentences. For ease of data retrieval, we also use <s></s> tags to delimit sentence segments. Table 1 tells us that 1) SS length in CTC and OCT are shorter than their counterparts in EST, 2) SS in TCT are significantly longer than those in OCT, and 3) mean SS length of TCT is very similar to that of EST.

Obviously, results of S length count are different from those of SS length count, yet they are not in conflict. For example, a Chinese sentence can be longer than an English one, but the former may contain more segments than the latter. Put differently, Chinese sentences contain more but shorter clauses or phrases than English sentences do, and this can explain why English, compared with Chinese, is shorter sententially but longer segmentally.

[1] A segment can be a clause or a phrase, with commas, semi-colons, colons being delimiters.

The greater sentence length of TCT supports explicitation in translation, but the greater SS length of TCT (compared with OCT) does not support normalization, because it is more like EST than OCT in this respect.

5. POS Distribution

5.1 Statistical Results

POS (part of speech) distribution partially reflects typological features of a language. For POS distribution analysis of Chinese and English, we use ICTCLAS and CLAWS to tag CTC/OCT and EST. In addition, to ensure the representativeness of normality, we single out literature subcorpora for analysis.

Table 2 POS Distribution in OCT, TCT and EST (lit.)

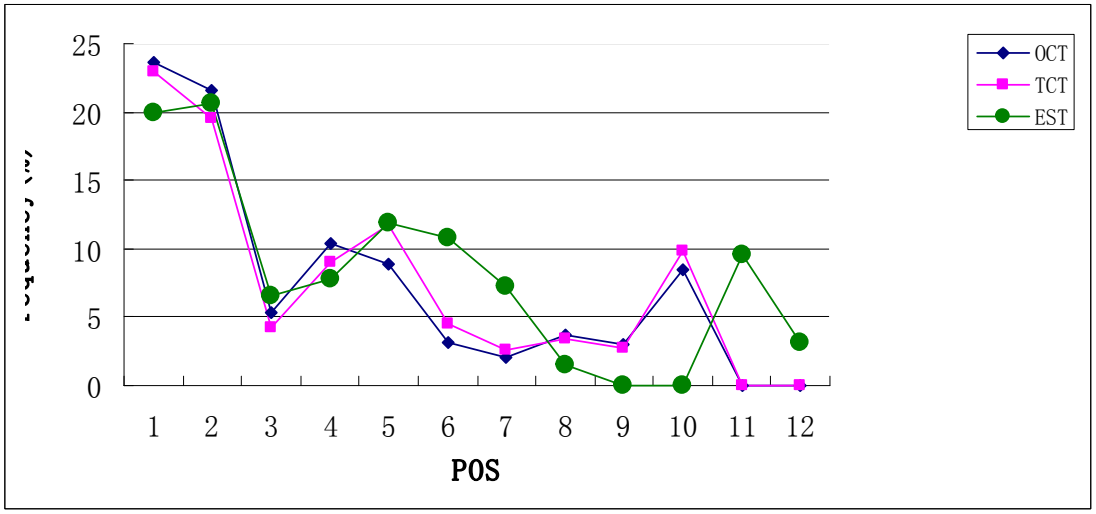
Distribution / POS		Original Chinese (lit)		Translational Chinese (lit)		Original English (lit)	
		Number	Frequency	Number	Frequency	Number	Frequency
1	**Verbs**	110391	23.64	133762	22.93	108340	19.88
2	**Nouns**	100827	21.59	113823	19.52	112536	20.65
3	**Adjectives**	24948	5.34	24672	4.23	35846	6.58
4	**Adverbs**	48676	10.42	52266	8.96	42065	7.72
5	**Pronouns**	41259	8.83	68859	11.81	64433	11.82
6	**Prepositions**	14536	3.11	25932	4.45	58743	10.78
7	**Conjunctions**	9687	2.07	15252	2.61	39304	7.21
8	**Numerals**	17322	3.71	20174	3.45	8463	1.5
9	**Classifiers (Ch)**	14209	3.04	16337	2.80	0	0
10	**Particles (Ch)**	39370	8.51	57372	9.84	0	0
11	**Articles (En)**	0	0	0	0	52325	9.60
12	**Determiners(En)**	0	0	0	0	17132	3.14
Total		421225	90.26	528449	90.6	539187	98.88

As is illustrated in Table 2, EST has a lower frequency of verb use, about 4% lower than OCT. This difference confirms our belief that English is prominently ‘nominal’ while Chinese is more ‘verbal’ (Si 2002: 55-58; Shao 2005: 24).

In addition, Table 2 displays difference in the use of pronouns, prepositions and

conjunctions. On the one hand, TCT is very similar to OET but quite different from OCT in the use of pronouns, which suggests TCT has undergone interference from EST; on the other hand, TCT uses much fewer prepositions and conjunctions than EST (their source texts), which means TCT seems to have gone through less interference from EST in the use of prepositions and conjunctions.

Figure 1 A broken line graph of POS distribution



Numbers along x-axis: 1=V; 2=N; 3=AJ; 4=AD; 5=PR; 6=CON; 7=PP; 8=NUM; 9=CLA; 10=PAR; 11=ART(En); 12=DET(En)

Figure 1 shows that TCT and OCT are very similar in line shape, but they are quite different from OEC. That suggests, as far as POS distribution is concerned, the difference between TCT and OCT is smaller than that between TCT and EST, and translational Chinese largely conforms to the normality of Chinese.

In sum, compared with OCT, TCT uses fewer verbs, nouns, adjectives and adverbs, but more pronouns, prepositions and conjunctions. In line with these facts, we have good reason to assume that POS distribution in TCT has received interference from English, the source language.

The analysis made so far shows TCT and OCT are similar in POS distribution, but it is still uncertain if this is true for the distribution of specific lexical items. For a close look at the

behaviors of specific lexical items, we employ Keywords tool in WordSmith 4.0 to extract words that have statistically significant difference in frequency in TCT and OCT (the reference corpus). In what follows, we observe content words in Section 5.2, and function words in Section 5.3.

5.2 Content Words

5.2.1 Nouns

We find that TCT and OCT are different in using nominal expressions. Lexically, the frequency of some words like 上帝(*God*) , 绅士(*gentleman*), 牧师(*priest*) are unusually high in TCT in comparison with OCT, while OCT marks an higher frequency in the use of nouns like当差(*lower official or servant in ancient China*), 洋车(*rickshaw*), 饺子(*jiaozi*), 表姐(*a daughter of father's sister or of mother's brother or sister*, who is older than oneself), 姑奶奶(*sister of one's paternal grandfather*, sometimes refers to speaker herself, arrogantly), 旗袍(*chi-pao*), 夜壶(*chamber pot*), to name just a few.

Morphologically, morphemes like兄(*brother*), 时(*time*), 堂(*woman, mother*), 氏(*surname*), 帖(*document or notes*), 斋(*studio*), etc. to form the words like令堂(your mother),午时(at midday) and白塔寺(Baita Temple) are almost only found in OCT.

Apparently, the difference in using nouns is attributed largely to cultural and social differences between the two language communities.

5.2.2 Verbs

In TCT, the following verbs are used with unusually high frequency: a) aspectual verbs: 开始(*begin*), 结束(*finish*); b) 'happen' verbs: 发生(*happen*), 产生(*produce*); c) 'find' verbs: 表现(*represent*), 发现(*discover*); d) causative verbs: 让(*let, make*); e) 'judge' verbs: 认为(*think*), 相信(*believe*), 感觉(*feel*); f) psychological verbs: 害怕(*afraid*), 怀疑(*doubt*); g) others: 具有(*have*),

存在(*exist, there be*). Certainly, these words also occur in OCT, but their frequency is remarkably lower. The major reason for this difference is that their counterparts (verbs, prepositional phrases, and adjectives) are used with usually high frequency in EST.

We also find OCT uses more monosyllabic verbs than TCT. The frequency of verbs such as 凑(*gather*), 搁(*place*), 甭(*don't*), 傍(*depend on*), 嚷(*shout*), 嫌(*dislike*), 吵(*quarrel*), 捧(*hold in both hands*), 混(*mix, make trouble, lurk*), 怔(*daze*) is usually high in TCT compared to OCT, which tells us that, from a monosyllable-idiomaticity correlation perspective, TCT is less idiomatic than OCT.

5.2.3 Adjectives

As keywords extraction shows, OCT frequently uses more monosyllabic adjectives than TCT, which are 脆(*crisp*), 高(*high*), 贵*(expensive*), 好(*good*), 红(*red*), 厚(*thick*), 慌*(nervous*), 紧(*tight*), 老(*old*), 俏(*charming*), 小(*small*), 饱(*full*), 苦(*bitter*), etc.. In contrast, TCT only frequently uses a tiny number of adjectives like大(*big*), 久(*long*), 多(*many*),快(*fast*), but it frequently uses more disyllabic adjectives. This is another evidence of less idiomaticity in language use in TCT.

5.2.4 Locative Particle

In OCT, only 里(*inside*), 外边(*outside*), 内(*within*) are more frequently used in comparison with TCT. In TCT, however, there are more such high-frequency locative particles, such as 以前(*before*), 之前(*before*), 之间(*between*), 之后(*after*), 之中(*in*), 之外(*outside*), 周围(*around*). The unusual frequency of locative particles in TCT results from interference from EST, for the latter frequently uses prepositions such as *before, after, between, in, under, near, around*, etc., and this high frequency in use is reflected in rather frequent use of locative particles in Chinese

translations.

5.2.5 Adverbs

In TCT, some time-related adverbs, such as 正(*in the process of*), 已(*already*), 已经(*already*), 一直(*always*), are much frequently found to co-occur with aspectual markers (ZHE, LE, GUO) to unify their temporal features. The reason for this is that TCT has a strong tendency to explicitate by lexical means the perfective and imperfective senses inherent respectively in *have v-en* and *have been v-ing* constructions.

Modal adverbs express speaker's attitude towards a proposition. In TCT, some modal adverbs are used with a strikingly high frequency, which are必须(*must*), 或许(*perhaps*), 竟然(*actually*), 大约(*about*), 如此(*so*), etc. OCT, in contrast, frequently uses other words to perform the function, such as 得（*have to*），兴许(*perhaps*)，原来(*actually*)，却(*actually*)，来（*about*），这么(*so*). So, the difference is not only in frequency, but also in word choice: Modal adverbs used in TCT seem to be more formal, far from being spoken and spontaneous as those used in OCT.

5.3 Function Words

5.3.1 Pronominals

In OCT, the following pronominals[1] show an usually high frequency: 大家(we, us), 她们(*they, them*), 怎(*what*), 怎样(*what*), 这(*this*), 自己(*self*), etc.

In contrast, TCT uses the following pronominals in usually high frequency, such as 她(she/her), 他(*he/him*), 他们(*they/them*),它(*it*), 它们(*them*), 我(*i/me*), 我们(*we/us*), 那(*that*), 那儿

[1] In Chinese linguistic literature, pronouns belong to content words, however, they form a closed set, so we prefer to say they are function words.

(*there*), 那个(*that one*), 那时(*then*), 那种(*that kind of*), 这个(*this*), 这时(*at the time*), 这种(*this kind of*), 这样(*in this way*), 其他(*other*), 另(*the other*), 别的(*other*), 任何(*anyone*), 每个(*everyone*), 一切(*all*). Obviously, in contrast with OCT, TCT seems to have exaggerated the use of 1st and 3rd personal pronouns, some demonstrative pronouns and some classifiers. In sum, the use of pronominals contributes to the uniqueness of translational Chinese. For more analysis, see Section 6.2.

5.3.2 Conjunctions

A major difference between TCT and OCT is found in the use of conjunctions. The statistics show that there are 16 conjunctions occurring with usually high frequency in TCT, such as不过(*but*), 但(but), 但是(*but*), 尽管(*though*), 或者(*or*), 不仅*(not only*), 而且(*moreover*), 另外(*in addition*), 哪怕(*even if*), 即使(*even though*), 如果(*if*), 然后(*then*), 因此(*so*), 于是(*upon that*) and 和(*and*). In contrast, only 7 conjunctions are used frequently in OCT, which are可是(*but*), 并且(*and*), 况且(*besides*), 不但(*not only*), 所以(*so*), 假若(*if*), 愈……愈(the more…the more), etc. This difference in number and frequency only confirms our belief that logical relations are more implicit in Chinese than in English.

The POS distribution analysis made so far suggests many features characteristic of TCT can look to interference from EST for explanation. In short, compared with OCT, TCT uses more function words and more disyllabic words.

6. Compositionality

TCT may exaggerate the compositional potentiality of some morphemes or words in Chinese. The exaggeration can be observed in the following three ways.

6.1 Nominal Morphemes

Take for example the Chinese morpheme性(-*xing*； meaning *property*, similar to *–ness, -ity)*. It occurs 2.9 times per ten thousand words in original Chinese literary texts; in contrast, its frequency hits 5.2 in translated Chinese literary texts. Moreover, its diversity in composition increases in TCT. For example, there are 71 types of -*xing* combinations in translated Chinese literary texts, of which 42 are not used or rarely used in Chinese original literary texts, such as 独创性（creativity）, 决定性(decisiveness),可信性(reliability), 坚定性(firmness), 实质性(substantiality), 强制性(compulsiveness), etc. In the original Chinese literary texts, however, -*xing* is found more often to occur with monosyllabic words, forming words like 爽性(straightforwardness), 火性(bad temper), 牛性(obstinacy), 癖性(natural inclination), 韧性(tenacity), etc.

In translation, a translator consciously or subconsciously follow and imitate some features of source language (Kefei 2005). The active and diverse use of -*xing* morpheme in TCT is a good case in point. In TCT, -*xing* imitates the corresponding suffixes such as -*ity*, -*ness*, -*dom,* etc. in English, hence more active than it is in OCT. Interestingly, −*xing* is now a regular morpheme in Mandarin Chinese.

The phenomenon mentioned above is true for morphemes such as −力 (*force, ability*)， −度 (*degree, extent*) in TCT where they are very frequently used to form technical terms.

In a word, the high-frequency use of these morphemes in TCT suggests an imitation of their compositionality in English.

6.2 Composition of "Dem + Num + Cla"

The construction "demonstrative pronoun(Dem)+numeral(Num)+classifier(Cla)" occurs

more frequently in TCT than in OCT. Listed below are three constructions with such high frequency.

Table 3 Distributions of "Dem+ Num+ Cla" Constructions

	TCT		**OCT**	
	Frequency	MI	Frequency	MI
这 一	1010	0.035	334	0.019
这 种	491	0.09	39	0.012
这件事	123	0.017		

In OCT, the Dem-Num composition这一 (*zhe yi*; literally, *this one*) typically co-occurs with monosyllabic words such as 点 (point), 条 (item), 天 (day), 次 (time), 年 (year) ; but in TCT, the construction's compositional potentiality is dramatically enhanced in that it also frequently co-occurs with disyllabic words like 问题(problem), 条款(article, clause, item), 事实(fact), 目标(aim), 领域(field), 计划(plan), 过程(process) and 观点(viewpoint). This enhancement contributes partially to high-frequency use of disyllabic words in TCT.

Another case in point is 这种 (*zhe zhong*; literally, *this kind*), this Dem-Num phrase is weak in compositionality in OCT because it almost exclusively co-occurs with 人(*person*). Even so, the combination "这种-人" is far less frequently used in OCT than in TCT (39 vs. 295). In TCT, *zhe zhong* is extremely active, and it can yield diversified compositions, as can be seen in its composition with 药(*medicine*), 病(*disease*), 事(*matter*), 做法(*practice*), 想法(*idea*), 现象(*phenomenon*), 感觉(*feeling*), 情况(*situation*), 方式(*way*), 方法(*method*), and with many other nominals that are not frequently or typically used in OCT.

The search in GCEPC shows the frequent use of 这一 and 这种in TCT corresponds to the frequent use of articles and demonstrative (such as *the/ this/ that*) in EST.

For the same reason, other such phrases as一个(*an-individual*), 一件(*an-item*), 一位(*a-*

position), 一片(*a-slice*) occur more frequently in TCT than in OCT. The major reason for this difference is that indefinite articles in English strongly tend to be rendered into Chinese Num-Cla phrases.

It should be noted that many Num-Cla-N bundles used in TCT are not always the direct translations of NPs in EST; in fact, some of them are renderings of pronouns or demonstratives. For example, 这件事 in TCT might be an equivalent of *it*, *this* or *that*, but not necessarily *the matter* or *the thing*.

（3）a. **It** had happened at last.

这件 事 终于 发生 -了。
This matter eventually happen-LE(crs)

b. "Nobody knows about **this** but us?"

“除了 咱们，没 人 知道 **这件 事** 吧？”
“except us, no man know this matter-BA?”

c. "Humph! We'll see about **that**."

“嗯，**这件 事** 我们 得 管一管-了。”
“Er, **this matter** we have to take care of-LE(cr) ”

d. There must 'a' been an angel **there**.

这件 事 一定 有 个 高手 在-帮 你的 忙。
This matter must have one master-hand ZAI-help your busy-work

The examples above demonstrate changes of cohesive devices in E-C translation, i.e. a change from pronouns, demonstratives or demonstrative adverbs in English to NPs in Chinese version. These changes reflect the difference between Chinese and English: the former employs more lexical devices to realize textual coherence. However, the changes are not frequently found in TCT; in fact, compared with OCT, TCT resorts more to the use of pronominals to achieve

textual cohesion, which means translational Chinese is very similar to English source texts in the use of cohesive devices.

All in all, some grammatical devices typically used in English are prone to being imitated in E-C translation, which leads to the overuse of the compositionality of some morphemes, phrases and cohesive devices and which makes TCT less idiomatic than OCT.

6.3 Relative Fixedness of Some Expressions

It is argued that some phrasal discourse markers such as comment clauses *I think*, *it seems* and conversational routines (such as *thank you*), are to some degree lexicalized, because they are relatively fixed usages (*see* Brinton & Traugott 2005: 67). In TCT, we do find some lexical bundles that are rather frequently used. They might not be very 'Chinese', but they tend to be 'institutionalized' and become rather fixed. Below are two examples.

随着时间的推移

The expression随着时间的推移 (*lit.* along with time's running) denotes the passage of time. It is often used in Chinese, yet it is borrowed from English (perhaps an imitation of *with the passage of time*). What interests us here is the fact that the expression can be used in dealing with many similar expression in English, such as *as time went on, moment by moment, over time, as time drifted along, with a long-term time horizon, in the course of time, as time went by*, or even *eventually.*

Actually, the expression is now a regular expression in Chinese, and is even more popular than the very 'Chinese' expressions like光阴荏苒(*time elapses quickly*),日复一日(*day after day*), 岁月流转(*with the passage of time*). Perhaps it is due to translation that the expression 随着时间的推移 becomes an expression frequently used in Chinese.

是(不)可能(的)

The expression是(不)可能(的) (lit. is (not) possible) conveys speaker's attitude toward a proposition. In TCT, this frequently-used expression is an equivalent to many expressions in EST. Below is a list of the possible equivalents.

Table 4 Expressions equivalent to 是(不)可能(的)

English Equivalents	POSSIBLE	CAN	LIKELY	WILL	Might	Probable	Incapable	No Equivalent	Total
Frequency	41	10	5	2	2	1	1	3	67

In OCT, the expression是(不)可能(的) appears only 17 times, all of which being found in political texts; in addition, it is usually put at the end of a sentence, hence not used frequently and diversely. In TCT, its use is diversified because it can appear at different positions in a clause (quite like the use of 'possibility' expressions in English).

What's more, in TCT, the load capacity of 是不可能[…]的is expanded. As in the example below, where the capacity of […] is expanded to 34 Chinese characters (in Chinese version, parts in italics are equivalent to the part following "the impossibility that" in English):

(4) […], but had long since recognized **the impossibility that** *any mission of divine and mysterious truth should be confided to a woman stained with sin, bowed down with shame, or even burdened with a life-long sorrow.*

[…]，但从那以后，她早已承认了：*任何上界的神秘真理的使命***是不可能***委托给一个为罪孽所玷污、为耻辱所压倒或者甚至为终生的忧愁而沉闷的女人***的**。

Through N-gram search within TCT, we find many other expressions behaving like the above two expressions, which are 目的是为了(*for the purpose of*), 在某种程度/意义上(*to some*

degree, in a sense), 是必要的(*it is necessary that*), 一遍又一遍(*time and again*), 很久很久以前(*long, long ago*), 更确切地说(*precisely*), 一般情况下(*generally*).

The fixedness of expressions in TCT suggests that once an expression enters into the target language through translation, it might become relatively fixed and frequently used in translational language or even grow into a popular expression in target language. TCT has a stronger tendency to use relatively fixed expressions to deal with diverse expressions (with same or similar functions) in EST, which can serve as a support for lexical simplification, but at the same time a denial of normalization in translation universals.

7. Conclusion

Translational Chinese has the following features: 1) TCT uses fewer monosyllabic words than OCT does; 2) TCT tends to expand the normal load capacity of some Chinese constructions, which leads to longer sentence segments; 3) compared with OCT, TCT uses more function words; 4) TCT can change or expand the compositionality of some words or morphemes in Chinese.

The features mentioned-above do not fully support the translation universals.

Firstly, TCT use more types and longer segments than OCT. This does not support lexical and syntactic simplification;

Secondly, explicitation in TCT runs in parallel with implicitation, such as the implicitation of logical relations and co-reference devices. In this sense, explicitation is a relative notion. As far as English-Chinese translation is concerned, TCT is more explicit than OCT, but more implicit than EST. This relativity suggests explicitation and implicitation co-exist in any translation pair; it is not always unidirectional.

Thirdly, TCT exaggerates the compositional potentiality of some morphemes and words in Chinese, and it has expanded some Chinese constructions' load capacity. Considering this, TCT does not fully support normalization.

It can be concluded that the so-called translation universals might be a shifting phenomenon between specific languages or just some features in local translated discourse. It is by no means the only phenomenon applicable to all translational languages.

References

1. Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications". In Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair*. Amsterdam, Philadelphia: John Benjamins, 233-250.
2. Baker, Mona. 1998a. Réexplorer la langue de la traduction: une approche par corpus. *Meta,* 43(4): 480-485.
3. Baker, Mona. 1998b. *Routledge Encyclopedia of Translation Studies*. London:Routledge.
4. Brinton, Laurel J. & Elizabeth Closs Traugott. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
5. Chen, Hsin-His. 1994. The Contextual Analysis of Chinese Sentences with Punctuation Marks. *Literary and Linguistic Computing,* 9(4):281-289.
6. Hansen S. & Teich, E. 2001. Multi-layer analysis of translation corpora: methodological issues and practical implications. In D. Cristea, N. Ide, D. Marcu and M. Poesio (eds.), *Proceedings of EUROLAN 2001 Workshop on Multi-layer Corpus-based Analysis*. Iasi, Romania, 44-55.
7. Hopkinson, Chris. 2007. Factors in Linguistic Interference: A Case of Study in Translation. *Skase Journal of Translation and Interpretation*, 2(1):13-23.
8. Huang, Libo & Wang, Kefei. 2006. "Fanyi Pubianxing Yanjiu Fansi"[Reflections on the corpus-based studies of translation universals]. *Chinese Translators Journal,* 5: 36-40.
9. Ke Fei. 2003. "Hanyu Ba Zi Ju Tedian, Fenbu Ji Yingyi"[Chinese BA-construction and its English translation]. *Foreign Languages and Their Teaching*, 12: 1-5.
10. Ke Fei. 2005. "Fanyi Zhong De Yin He Xian"[Implicitation and explicitation in translation]. *Foreign Language Teaching and Research,* 4: 303-307.
11. Lü, Shuxiang (ed.). 1980. "Xiandai Hanyu Ba Bai Ci"[*Eight Hundred Words in Mandarin Chinese*]. Beijing: The Commercial Press.
12. Qin, Hongwu & Wang, Kefei. 2004. "Jiyu Yuliaoku De Fanyi Yuyan Fenxi"[Parallel corpora-based analysis of translationese]. *Modern Foreign Languages*, 1:44-52.

13. Santos, Diana. 1995. On grammatical translationese. In Kimmo Koskenniemi (ed.), *Short papers presented at the Tenth Scandinavian Conference on Computational Linguistics.* Helsinki: 59-66.
14. Shao, Zhihong. 2005. "Han Ying Duibi Fanyi Daolun"[*Chinese-English Contrastive Analysis and Translation*]. Shanghai: Huadong University of Technology Press.
15. Si, Guo. 2002. "Yi Dao Tanwei"[*Looking into Translation*]. Beijing: China Translation & Publishing Corporation.
16. Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
17. Westin, Ingrid. 2002. *Language and Computers: Studies in Practical Linguistics*. New York: Rodopi.
18. Wu, Ang & Huang, Libo. 2006. "Guanyu Fanyi Gongxing De Yanjiu" [On corpus-based studies of translation universals]. *Foreign Language Teaching and Research,* 5:296-302.
19. Yu, Guangzhong. 2002. "Yu Guangzhong Tan Fanyi" [*On Translation*]. Beijing: China Translation & Publishing Corporation.