

Lexical and grammatical properties of Translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective

RICHARD XIAO and GUANGRONG DAI

Abstract

Corpus-based Translation Studies focuses on translation as a product by comparing comparable corpora of translated and non-translated texts. A number of distinctive features of translations have been posited including, for example, explicitation, simplification, normalisation, levelling out, source language interference, and under-representation of target language unique items. Nevertheless, research of this area has until recently been confined largely to translational English and closely related European languages. If the features of translational language that have been reported on the basis of these languages are to be generalised as “translation universals”, the language pairs involved must not be restricted to English and closely related European languages. Clearly, evidence from a genetically distant language pair such as English and Chinese is arguably more convincing, if not indispensable. This article explores, in the broad context of translation universal research, lexical and grammatical properties of translational Chinese on the basis of two one-million-word balanced comparable corpora of translated and non-translated native Chinese texts. The findings of this empirical study of the properties of translational Chinese have enabled a reevaluation, from the perspective of translational Chinese, of largely English-based translation universal hypotheses.

Keywords: *corpus-based approach, translation universal, translational Chinese, lexical and grammatical properties*

1. Introduction

Textual studies that compare translated texts with source texts and originally composed target texts started in the early 1980s, which show that translational language as a type of mediated discourse has distinctive features that make it perceptibly different from comparable target language. For example, Duff (1981: 12) finds that a translated text often represents “a mixture of styles and languages”, or a “patchwork” made up of SL [source language] and TL [target language] elements, and consequently, translational language is labelled the “third language” that lies between the source language and the target language: “The world may be approached from a different angle and the information given may yet be the same” (ibid. 13). Frawley (1984: 168) also argues that translation “is essentially a third code which arises out of the bilateral consideration of the matrix and target codes: it is, in a sense, a sub-code of each of the codes involved.” Likewise, Blum-Kulka (1986: 19) puts forward the “explicitation hypothesis”, for the first time, on the basis of her investigation of shifts of cohesion and coherence in translation, which posits that explicitation is “inherent in the process of translation.”

Since corpus linguistics brought about a paradigmatic shift in Translation Studies in the early 1990s (cf. Laviosa 1998a), these observations have been supported by empirical evidence from a range of corpus-based studies of the features of translational language in relation to source and target languages, which are often referred to as “translation universals”. According to Baker (1993: 243), translation universals (TUs) refer to the features “which typically occur in translated texts rather than original utterances and which are not the result of interference from specific linguistic systems.” They are recurrent common properties of all translated texts, which are “almost the inevitable by-products of

the process of mediating between two languages rather than being the result of the interference of one language with another” (Laviosa 2002: 43). Hence, it is commonly assumed in Translation Studies that translated texts differ not only from their source texts but also from comparable native texts in the target language (cf. Hansen and Teich 2001:44). As McEnery and Xiao (2008) observe, translational language is at best an unrepresentative special variant of the target language.

Translation universal hypotheses have been a focus of research over the past two decades. During this period of time corpus-based studies have uncovered, mostly on the basis of contrastive analyses of translational English in comparison with native English, a number of common properties of translated texts, which are supposed to be universal features of all translational languages, e.g. explicitation, simplification, normalisation, source language (SL) interference, target language (TL) unique item under-representation, and levelling out (see Section 2 for a review).

However, while there is increasing consensus that translated texts are distinct from original writings in the target language, the TU hypotheses have also been a target of debate. The debates centre around two issues, the first of which relates to the appropriateness of making universal claims. For example, Pym (2008) flatly rejects the very concept of universals, while Tymoczko (1998), Malmkjær (2005) and House (2008) think it inconceivable to make universal claims about translation.¹ It has also been argued, however, that the main value of translation universals (or general “laws” of translation) lies in their “explanatory power”, “even though not necessarily under the title of ‘universals’” (Toury 2004: 29).

The second issue is whether the features of translations that have been posited on the basis of translational English are also generalisable to other translational languages, especially when translation involves languages that are genetically distinct. For example, Cheong’s (2006) study of English-Korean translation contradicts even the least controversial explicitation hypothesis. It has to be admitted that TU research has largely been Eurocentric, until very recently, in that existing evidence in support of the proposed TU hypotheses has mostly come from translational English and related European languages (cf. Xiao 2010). This is unsurprising given that the pioneering Translational English Corpus (TEC) has been the only publically available corpus of translational language that has provided an empirical basis for most of the prominent studies of translational English (e.g. Baker 1996; Kenny 1998, 2001; Laviosa 1998b, 2002; Olohan and Baker 2000; Olohan 2001, 2004), until more recently when the Corpus of Translated Finnish (CTF, see Mauranen 1998) and the ZJU Corpus of Translational Chinese (ZCTC, see Xiao, He and Yue 2010) were created. However, if the discussion of TUs in Translation Studies follows the general discussion of universals in language typology (cf. Mauranen 2008), it can reasonably be argued that genetically distant languages such as English and Chinese will provide evidence that is of critical importance and is more convincing than English and closely related Germanic and Romance languages if the features of translations that have been observed on the basis of translational English are to be generalised as “translation universals”.

Building on Xiao et al. (2010) and Xiao (2010, 2011, 2012), the present study will further explore the lexical and grammatical properties of translational Chinese on the basis of two balanced comparable corpora of native and translational Chinese, and reevaluate existing TU hypotheses in the face of evidence from translational Chinese, with the aim of addressing two research questions:

- 1) In comparison with comparable native Chinese writings, what lexical and grammatical properties do translated Chinese texts have in common?
- 2) What are the implications of the properties of translational Chinese for the English-based TU hypotheses?

The major corpus resources used in this study include the Lancaster Corpus of Mandarin Chinese (LCMC), which was created by following the FLOB corpus design (Hundt et al. 1998) to

represent native Chinese as used in the early 1990s for use in contrastive studies of English and Chinese (McEnery and Xiao 2004), and its translational counterpart, the ZJU Corpus of Translational Chinese (ZCTC). Like FLOB, both Chinese corpora comprise five hundred 2,000-word text chunks sampled from the same fifteen genres as indicated in Table 1, with each corpus totalling one million word tokens. The text samples included in both the native and the translational Chinese corpora are all taken from materials published in China in comparable sampling periods. English is the source language for 99% of the text samples in ZCTC.

Table 1. LCMC and ZCTC corpus design

Type	Register	Code	Genre	Samples	Proportion
Non-literary	News	A	Press reportage	44	8.8%
		B	Press editorials	27	5.4%
		C	Press reviews	17	3.4%
	General prose	D	Religious writing	17	3.4%
		E	Instructional writing	38	7.6%
		F	Popular lore	44	8.8%
		G	Biographies and essays	77	15.4%
		H	Reports & official documents	30	6%
	Academic prose	J	Academic writing	80	16%
	Literary	Fiction	K	General fiction	29
L			Mystery & detective fiction	24	4.8%
M			Science fiction	6	1.2%
N			Adventure fiction	29	5.8%
P			Romantic fiction	29	5.8%
R			Humour	9	1.8%
Total				500	100%

Following this introduction, Section 2 provides a brief review of the TU hypotheses that have been proposed in previous research. Sections 3 and 4 explore the lexical and grammatical properties of translational Chinese respectively, which is followed by a reevaluation of TU hypotheses from the perspective of translational Chinese (Section 5). Section 6 concludes the article by summarising the major research findings.

2. Translation universal hypotheses

According to Chesterman (2004: 39), translation universals are higher level generalisations of the common properties of translated texts, which can be either “universal differences between translations and their source texts, i.e. characteristics of the way in which translators process the source text” (i.e. S-universals), or “universal differences between translations and comparable non-translated texts, i.e. characteristics of the way translators use the target language” (i.e. T-universals). The convergence between corpus linguistics and Translation Studies has, since the early 1990s, greatly facilitated what Toury (1995) calls “product-oriented translation research”. Within this area of enquiry a number of translation universal hypotheses have been proposed such as explicitation, simplification, normalisation, SL interference, TL unique item under-representation, and levelling out. Explicitation and SL

interference are considered to be S-universals, whereas simplification, normalisation and TL unique item under-representation can be said to be T-universals, while levelling out can be either. This section will briefly review these key concepts so as to set the scene for the research presented in the following sections. Readers can refer to Xiao and Yue (2009) and Xiao (2010) for a fuller review of TU research.

2.1 *Explicitation*

In the prescriptive paradigm of Translation Studies, explicitation and implicitation are regarded as translation techniques or strategies that correspond to addition and subtraction (Nida 1964), with explicitation being defined as “the process of introducing information into the target language which is present only implicitly in the source language, but which can be derived from the context or the situation”, and implicitation as “the process of allowing the target language situation or context to define certain details which were explicit in the source language” (Vinay and Darbelnet 1958: 8; cited in Klaudy 2009: 80). The explicitation hypothesis is first put forward by Blum-Kulka (1986: 19):

The process of interpretation performed by the translator on the source text might lead to a TL text which is more redundant than the SL text. This redundancy can be expressed by a rise in the level of cohesive explicitness in the TL text. This argument can be stated as “*the explicitation hypothesis*”, which postulates an observed cohesive explicitness from SL to TL texts regardless of the increase traceable to the differences between the two linguistic and textual systems involved.

Given that Blum-Kulka’s investigation focuses on shifts of cohesion and coherence in translation, it is hardly surprising that she has defined explicitation in a narrow sense to refer exclusively to “cohesive explicitness”. In contrast, Baker (1996: 180) proposes a broad definition of explicitation, i.e. the tendency in translation to “spell things out rather than leave them implicit.” It follows that cohesive explicitness is merely one type of explicitation, which can also be realised at the semantic and grammatical levels (cf. also Mauranen 2008: 39).

Blum-Kulka (1986) appears to maintain a distinction between explicitation necessitated by cross-linguistic differences between the source and target languages and translational explicitation arising from the translation process itself (cf. Baumgarten, Meyer and Özçetin 2008). Klaudy (2009) proposes a more fine-grained classification of explicitation including four types. In addition to obligatory explicitation and translation-inherent explicitation, which respectively correspond to Blum-Kulka’s (1986) two types, optional explicitation results from different text building strategies and stylistic preferences between the source and target languages while pragmatic explicitation specifically relates to cultural differences explained by translators. As noted earlier, explicitation is essentially a type of S-universal in Chesterman’s (2004) terms. This means that explicated instances can be identified by comparing the source and target texts in a parallel corpus. When a parallel corpus approach is taken to investigate explicitation as an S-universal, it is of critical importance to take account of different kinds of explicitation because not all explicated instances are the optional choices open to the translator. It is also true to say that explicitation can be studied as a T-universal on the basis of monolingual comparable corpora composed of translated texts and comparable non-translated texts in the target language. This comparable corpus approach is taken in the present study to investigate explicitation in translational Chinese. In this approach, the distinction between different types of explicitation is of little relevance as different variants (namely translated versus non-translated) of the same language are compared and contrasted.

Despite some criticisms, e.g. House (2008: 10) who argues that “the quest for translation universals is in essence futile” and Becher (2010: 2) who makes a plea for abandoning “the dogma of

translation-inherent explicitation”, explicitation is probably the most studied and least controversial TU hypothesis that has been investigated up to the present time. In Section 5 we will reevaluate this hypothesis in the face of evidence from translational Chinese.

2.2 Simplification

Explicitation and simplification may overlap (Mauranen 2008: 41), with explicitation leading to simplification in that a more explicit message is expected to be easier to read. The simplification hypothesis concerns the “tendency to simplify the language used in translation” (Baker 1996: 181-182). Simplification has been observed at different levels. For example, simplification at the lexical level has been defined as “making do with less words” (Blum-Kulka and Levenston 1983: 119). Simplification also involves using informal, colloquial and modern lexis to translate formal, literate and archaic words in the source text (Vanderauwera 1985) and showing a preference for high-frequency words, lower lexical density, greater repetition of commonly used words, and less lexical variability (Laviosa 1998b, 2002). Syntactical simplification occurs when syntactic complexity is reduced by replacing non-finite clauses with finite clauses (Vanderauwera 1985), and when stronger punctuations are used to split lengthy and complex sentences in the source texts into short simpler structures in translated texts (Malmkjær 1997). Stylistic simplification relates to the translational practice of “replacing elaborate phraseology with shorter collocations, reducing or omitting repetitions and redundant information, shortening overlong circumlocutions and leaving out modifying phrases and words” (Laviosa 1998b: 289).

The simplification hypothesis is more controversial than the explicitation hypothesis. Laviosa-Braithwaite (1996), one of the first corpus studies of TU hypotheses, cautioned that earlier studies that had put forward the simplification hypothesis had failed to provide adequate evidence. The hypothesis has since been contested by a number of studies that have reported on more complicated linguistic features in translated texts than in non-translated texts in the target language, e.g. greater mean sentence length (Laviosa 1998b), more untypical collocations (Mauranen 2000), and more frequent use of modifiers (Jantunen 2004).

2.3 Normalisation

According to Baker (1996: 183), normalisation refers to “the tendency to conform to patterns and practices that are typical of the target language, even to the point of exaggeration”. This is compatible with Toury’s (1995: 268) law of growing standardisation, which states that “in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire.” Examples of normalisation evidenced by empirical studies include adapting odd punctuation marks in the source language to the target language norm, replacing metaphors or idioms in the source language with canonical ones that are functionally similar in the target language, overusing target language clichés, and using standardised target language to translate dialects in literary source texts (e.g. Baker 1996; May 1997; Mauranen 2008). In addition, Olohan (2004) also takes less variation in colour synonyms in translation as evidence in support of her claim for normalisation, while Kenny’s (2001) investigation of “creative hapax legomena” (i.e. words that occur only once other than those technical terms and non-standard orthographic variations and so on) supports normalisation as well as translators’ creativity.

Normalisation is also a very debatable TU hypothesis. Mauranen (2000) finds that untypical collocations are more frequently used in translated than comparable native texts. Toury (1995: 208) himself, while formulating his law of growing standardisation, concedes to “the well-documented fact

that in translation, linguistic forms and structures often occur which are rarely, or perhaps even never encountered in utterances originally composed in the target language.”

2.4 Source language interference

In addition to the law of growing standardisation, Toury (1995) puts forward the law of interference, which concerns the influence of the source language upon the translated text. Importantly, Toury observes that the status of the source language in the target culture influences the operation of these two laws, so that when the source language enjoys a high status, interference is more likely to occur. In contrast, when the source language has a low status, standardisation or normalisation is more likely to occur.

Interference is an S-universal in Chesterman’s (2004) terms, as it is a result of source text features being carried over into the target text. Teich (2003: 145) observes that “in a translation into a given target language (TL), the translation may be oriented more towards the source language (SL), i.e. the SL shines through”. For example, Teich’s (2003: 207) study indicates that in translation between English and German, the target texts in both directions of translation represent a mixture of normalisation and source language interference.

Indeed, source language interference is prevalent in translation. As Toury (1979: 226) notes, “virtually no translation is completely devoid of formal equivalents, i.e., of manifestations of interlanguage”, because interference may arise from any aspect of the make-up of the source text. In addition to the status of the source language and the attitude towards interference (Laviosa 2009: 307), the extent of interference is also expected to be negatively correlated with the translator’s competence and experience.

2.5 Unique item under-representation

As a T-universal, the hypothesis of TL unique item under-representation was put forward by Tirkkonen-Condit (2002), which was inspired by Reiss (1971) who claimed that “translations may not fully exploit the linguistic resources of the target language” (Tirkkonen-Condit 2002: 208). Tirkkonen-Condit (2004: 177) defines unique items in the target language as linguistic elements that “lack straightforward linguistic counterparts in other languages”, which tend to be, but are not necessarily, untranslatable. She accounts for TL unique item under-representation from the cognitive perspective: under-representation of TL unique items is due to the under-representation of such items in the translator’s mental lexicon because the source text has no such items to trigger them during the translation process (cf. Malmkjær 2005: 18). Nevertheless, under-representation can be equally explained by Toury’s law of interference, in our view. As translation is triggered by linguistic elements in the source text, if the source language does not have the linguistic features unique to the target language, the natural expectation is that such elements will be under-represented in the translated texts.

The evidence supporting the TL unique item under-representation hypothesis has come from studies of lexis, syntax, and syntax- pragmatics interface (Mauranen 2008: 42). In spite of some scepticism (e.g. Chesterman 2007), unique item under-representation has been considered as “an excellent candidate for the status of a universal” because it “receives a cognitive explanation” and is supported by evidence from two unrelated languages, i.e. Swedish and Finnish (Malmkjær 2005: 18).

2.6 Levelling out

Levelling out relates to the “tendency of translated text to gravitate towards the centre of a continuum” (Baker 1996: 184). It is also called “convergence”, which means the “relatively higher level of

homogeneity of translated texts with regard to their own scores on given measures of universal features” (Laviosa 2002: 72), or less variance in textual features in translated than native texts (Olohan 2004: 100). The evidence that Baker (1996: 184) gives in support of the levelling out hypothesis is her observation that “the individual texts in an English translation corpus are more like each other in terms such as lexical density, type-token ratio and mean sentence length than the individual texts in a comparable corpus of original English”; she also cites Shlesinger’s (1989) finding about interpreting to show that the distinctions between oral and written translations are reduced.

Levelling out is probably the least studied TU hypothesis, with little empirical research that has been undertaken to verify its validity. Pym (2008) argues against levelling out as a translational feature because it entails that any extreme explicitation, simplification and normalisation would not occur as they mean straying from the centre of a continuum. However, it is equally or probably more justified to argue that simplification, explicitation and so on do not have to go to the extreme in order for them to be considered as common features of translations. They are rather in relative terms, i.e. relative to the source language (in the case of S-universals) or the target language (in the case of T-universals). It is the hypothesised common features discussed in Sections 2.1 – 2.5 above,² if validated, that make translated texts more homogeneous and convergent towards each other.

3. Lexical properties of translational Chinese

In the ensuing sections we explore a range of lexical and grammatical properties of translational Chinese mainly on the basis of the ZCTC and LCMC data. These corpora have also been used in Xiao (2010, 2011) and have produced interesting preliminary results on translated Chinese texts in relation to comparable original Chinese texts in terms of general statistics, such as lexical density, information load, high frequency words, mean sentence length, word clusters as well as lexical and grammatical properties including the use of reformulation markers and conjunctions. Building on Xiao (2010, 2011), the present study will further explore the properties of translational Chinese, focusing on the lexical level in this section and on the grammatical level in Section 4. In what follows the following lexical properties will be investigated: word frequency and word length (Section 3.1), keywords (Section 3.2), word class distribution (Section 3.3), as well as the use of pronouns and prepositions (Section 3.4), idioms (Section 3.5), and major types of punctuation (Section 3.6).

3.1 Word frequency and word length

The contrastive analysis of LCMC and ZCTC presented in Xiao (2010) demonstrates that in terms of lexical density defined as the proportion of content words in total words (Stubbs 1996; Laviosa 1998b), native Chinese displays a significantly higher overall score than translational Chinese (66.93% vs. 61.59%, $t = -4.94$ for 28 d.f., $p < 0.001$), with 14 out of 15 genres covered in the corpora showing a significant difference, suggesting that native Chinese has a greater informational load than translational Chinese, while translational Chinese shows a higher proportion of function words. The standardised type-token ratios, on the other hand, do not differ significantly between the two corpora (46.58 vs. 45.73, $t = -0.573$ for 28 d.f., $p = 0.571$), with marginal differences in most of the genres, suggesting that native Chinese and translational Chinese do not differ much in lexical variability. In addition, in comparison with native Chinese, translational Chinese also has a greater accumulated proportion of high frequency words which are defined as words with a minimum percentage of 0.1% of the total corpus (35.70% vs. 40.47% respectively for native and translational Chinese), a higher ratio between high- and low-frequency word tokens (0.5659 vs. 0.6988), and a higher repetition rate of high frequency words (2,870.37 vs. 3,154.37). These results show that Laviosa’s (1998b) observations of the core patterns of lexical use in translational English are also supported by evidence from Chinese.

This section will further analyse the distributions of high- and low-frequency words as well as mean word length as measured by the number of syllables.

Figure 1 illustrates the distribution of high-frequency words with a minimum percentage of 0.5%, 0.1%, 0.07%, 0.05%, 0.03%, 0.02% and 0.01% of the respective corpus. As can be seen from Figure 1, the numbers of ultra-high frequency (greater than 0.05%) words are very similar in LCMC and ZCTC as these words are the core vocabulary items, including most function words, which are commonly used in both native and translational Chinese. We noted earlier that in terms of word tokens, translational Chinese has a higher ratio between high- and low-frequency words, but as Figure 1 shows, in terms of word types, native Chinese uses more high-frequency words, particularly sub-high frequency words (i.e. between 0.05% and 0.01%).

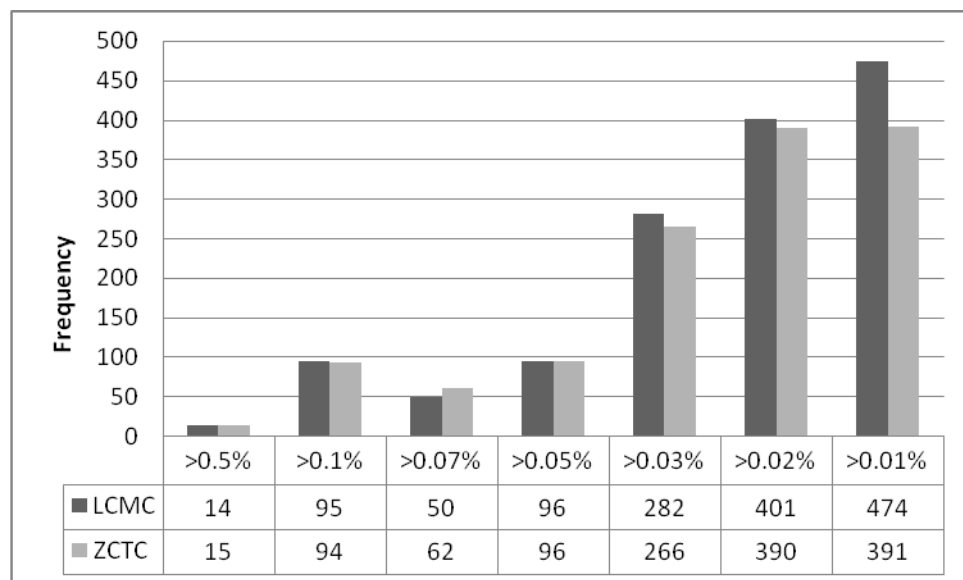


Figure 1. High frequency words in LCMC and ZCTC

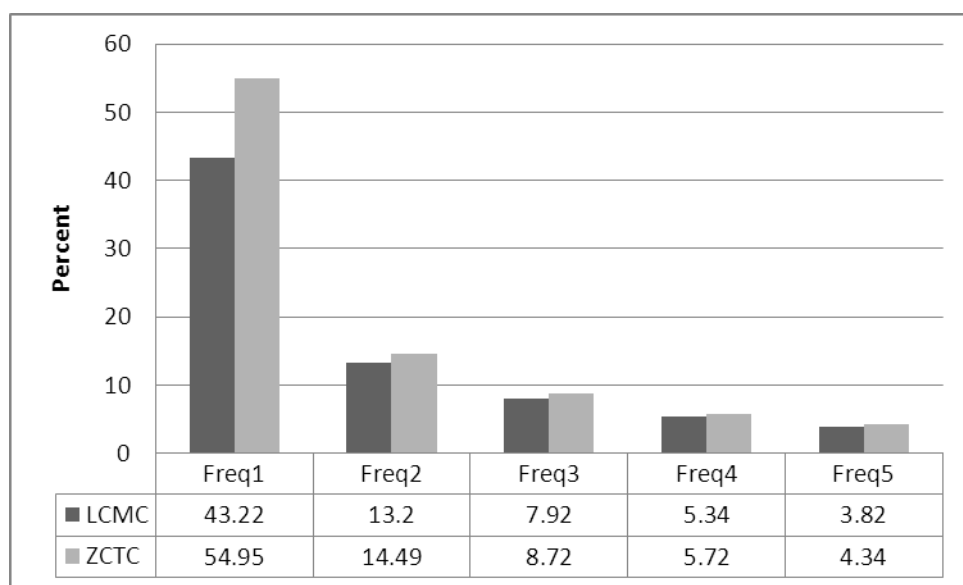


Figure 2. Low frequency words in LCMC and ZCTC

This tendency is reversed in low-frequency words. Figure 2 shows the distribution of words with a frequency of 1-5 in the native and translational corpora, where the percentage refers to the

proportion of the types of words with a particular frequency in total word types. It is clear from the figure that all of these low-frequency words display a high proportion in translational than native Chinese, and the lower the frequency is, the more marked the contrast is between the two corpora.

The discussions above suggest that translational Chinese, while demonstrating a preference for ultra-high-frequency words, also makes less frequent use of sub-high-frequency words but instead makes more frequent use of low frequency words. Hence, in terms of vocabulary use in general, translational Chinese does not necessarily demonstrate a simple tendency for simplification. In addition, as illustrated in Figure 3, the mean word length in translational Chinese is marginally greater than in native Chinese (1.59 vs. 1.57, a statistically insignificant difference), which is true in both non-literary (1.63 vs. 1.61) and literary (1.47 vs. 1.42) texts, with an even more marked contrast between native and translational Chinese in literary texts possibly because literary genres contain more proper nouns such as personal names and place names, which are longer than similar words in native Chinese.

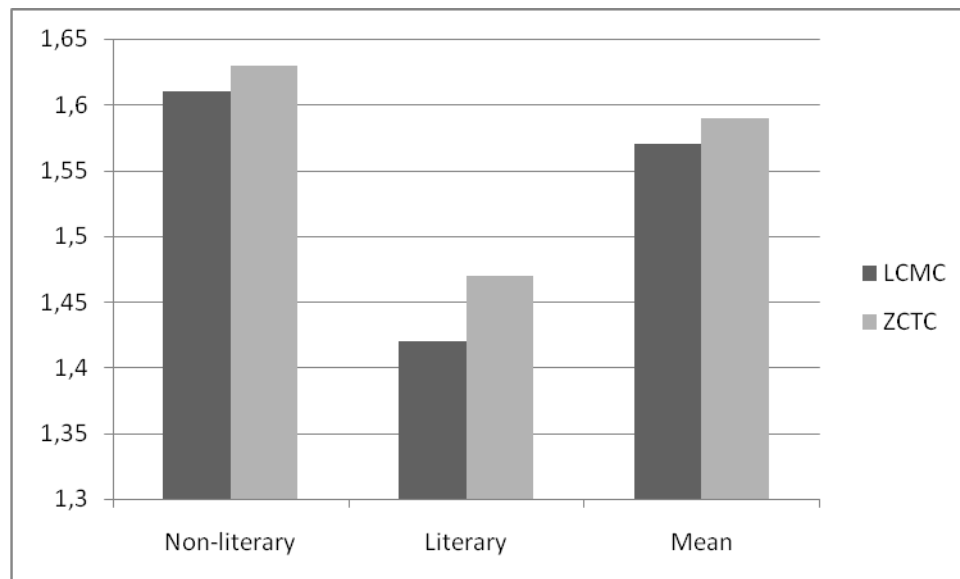


Figure 3. Mean word length in LCMC and ZCTC

Table 2. Proportions of words of various lengths in LCMC and ZCTC

Length	Non-literary texts		Literary texts		Mean score	
	LCMC	ZCTC	LCMC	ZCTC	LCMC	ZCTC
1 syllable	46.76	46.06	62.45	58.61	50.68	49.14
2 syllables	47.65	48.04	34.02	37.5	44.25	45.45
3 syllables	3.60	3.90	2.54	2.73	3.34	3.62
4 syllables	1.59	1.43	0.91	1.05	1.42	1.34
5 syllables	0.31	0.37	0.06	0.09	0.25	0.30
6+ syllables	0.09	0.19	0.01	0.02	0.07	0.15

Table 2 shows the distribution of words of various lengths in LCMC and ZCTC across two broad categories, namely literary and non-literary texts, and their mean scores. As the subcorpora are of different sizes, relative frequencies in the form of percentages will be compared. As can be seen from Table 2, no matter whether native and translational corpora are taken as a whole, or the two broad text categories are considered separately, monosyllabic and quadrisyllabic words are generally more frequent (except for quadrisyllabic words in literary texts) in native Chinese. A keyword analysis similar to that undertaken in Section 3.2 below suggests that monosyllabic words are more frequent in LCMC because native Chinese makes more frequent use of Chinese surnames, which are typically

monosyllabic, as well as high-frequency monosyllabic words such as 元 *yuan* ‘Chinese currency unit’ and 党 *dang* ‘(Communist) Party’, though many monosyllabic function words are more frequently used in the translational corpus, e.g. the structural auxiliary 的 *de* and personal pronouns 你 *ni* ‘you’, 我 *wo* ‘I, me’ and 她 *ta* ‘she, her’, which are all negative keywords in LCMC in relation to ZCTC (see Section 3.2 below for a discussion of keywords in ZCTC). Quadrisyllabic words are more frequently used in LCMC because non-literary texts in native Chinese tend to make significantly more frequent use of idioms (see Section 3.5 below), which are typically of the four-character-mould. In contrast, disyllabic and trisyllabic words are more frequent in translated texts. The translational tendency for long words is particularly marked in words containing five or more syllables, though these words per se are infrequent in both native and translational Chinese.

3.2. Keywords

Keywords are a powerful corpus linguistic tool that proves useful in content analysis as well as in stylistic research (Xiao and McEnery 2005). In this study, LCMC is used directly as the reference corpus in analysing keywords in ZCTC as the keywords extracted in this way are more characteristic of translational Chinese. Both (positive) keywords (i.e. those that are exceptionally frequent in the translational corpus) and negative keywords (i.e. those that are significantly infrequent in the translational corpus) will be included in keyword analysis.

Of the 100 most significant keywords in ZCTC, nouns that are often mentioned in translated texts take up the largest part (45 in total, e.g. 公司 *gongsi* ‘company’, 美国 *Meiguo* ‘the US’, 美元 *meiyuan* ‘US dollar’, 英国 *Yingguo* ‘Britain’, 全球 *quanqiu* ‘global’, 网络 *wangluo* ‘network’, 克林顿 *Kelindun* ‘Clinton’, 俄罗斯 *Eluosi* ‘Russia’, and 关税 *guanshui* ‘tariff’), followed by English character strings (18) in addition to some verbs (6), adverbs (5) and numbers (4), which are all required to express the intended ideational meanings. It can be said that these words appear on the keyword list out of the need for expressing the contents. As this study aims to reveal the characteristics of the language used in translational Chinese rather than comparing the contents expressed by native and translational languages, we will focus on the function words on the keyword list, which are deemed to be more useful than content words in studying the language form.

Table 3. Key function words in ZCTC

Category	Keywords
Pronoun	他 <i>ta</i> ‘he, him’; 我 <i>wo</i> ‘I, me’; 他们 <i>tamen</i> ‘they, them’; 她 <i>ta</i> ‘she, her’; 这些 <i>zhexie</i> ‘these’; 该 <i>gai</i> ‘this’; 它 <i>ta</i> ‘it’; 所有 <i>suoyou</i> ‘all’; 你 <i>ni</i> ‘you’; 那些 <i>naxie</i> ‘those’; 任何 <i>renhe</i> ‘any’; 我们 <i>women</i> ‘we, us’; 它们 <i>tamen</i> ‘they, them’; 其他 <i>qita</i> ‘other’; 这种 <i>zhezong</i> ‘this kind’
Conjunction	但是 <i>danshi</i> ‘but’; 如果 <i>ruguo</i> ‘if’; 因为 <i>yinwei</i> ‘because’; 或 <i>huo</i> ‘or’
Auxiliary	的 <i>de</i> ‘attributive modifier marker’; 所 <i>suo</i> ‘nominal marker’
Preposition	在 <i>zai</i> ‘at, in’

Table 3 lists the function words which are keywords in ZCTC. Function words on the ZCTC keyword list indicate that translational Chinese makes significantly more frequent use of function words, which is in line with our observation in Section 3.1. As can be seen from Table 3, pronouns (particularly personal pronouns and demonstrative pronouns) are more commonly used in translational Chinese. In terms of discourse analysis personal pronouns and demonstrative pronouns help to achieve textual cohesion and coherence while conjunctions explicate the logical relationships between clauses.

The most prominent function word on the ZCTC keyword list, which is also the third most significant keyword (the first two being content words 公司 *gongsi* ‘company’ and 美国 *Meiguo* ‘the US’), is the structural auxiliary 的 *de*, which is the attributive modifier marker in Chinese. This suggests that nouns in translational Chinese are more likely to take attributive modifiers, thus explaining why average sentence segments in translational Chinese are longer than in native Chinese (see Section 4.1 below).

The negative keywords in the translational corpus are mostly related to the contents expressed in native Chinese, e.g. 我国 *woguo* ‘our country’, 元 *yuan* ‘Chinese currency unit’, 同志 *tongzhi* ‘comrade’, 干部 *ganbu* ‘cadre’, 社会主义 *shehuizhuyi* ‘socialism’, 建设 *jianshe* ‘construction’, and 群众 *qunzhong* ‘masses’, which rarely appear in translated texts. However, there are also a number of negative keywords that are of interest, e.g. auxiliaries 了 *-le* ‘a perfective aspect marker’, 着 *-zhe* ‘an imperfective aspect marker’, 之 *zhi* ‘a structural auxiliary’ and 等 *deng* ‘etc.’; pronouns 咱们 *zanmen* ‘we’ and 啥 *sha* ‘what’; verbs 有 *you* ‘have’, 搞 *gao* ‘do’ and 办 *ban* ‘do’; as well as conjunctions 不但 *budan* ‘not only’ and 虽 *sui* ‘though’. These negative keywords reveal several tendencies in translated texts. First, translational Chinese tends to under-use aspect markers such as 了 *-le* and 着 *-zhe* (see Section 4.5); second, translational Chinese tends to avoid colloquial (e.g. 咱们 *zanmen* ‘we’), dialectal (啥 *sha* ‘what’), and archaic (之 *zhi*) lexis; third, translational Chinese may under-use light verbs (e.g. 搞 *gao* ‘do’, 办 *ban* ‘do’); finally, translational Chinese displays a lexical preference that is different from native Chinese: while native Chinese may use near synonyms alternatively for lexical variation (e.g. 不仅 *bujin* ‘not only’ vs. 不但 *budan* ‘not only’; 虽然 *suiran* ‘though’ vs. 虽 *sui* ‘though’), translational Chinese tends to repeat commonly used words (e.g. 不仅 *bujin* ‘not only’ and 虽然 *suiran* ‘though’).

3.3 Word class distribution

The present section explores the patterns of distribution of content words and function words. Figures 4 and 5 respectively compare the distribution of content word classes and major function word classes in LCMC and ZCTC.³ As the two corpora are of roughly equal size, raw frequency counts are directly comparable. As can be seen from Figure 4, all of the four content word classes are more frequent in native Chinese though the difference in adverbs is not as marked as in nouns, verbs and adjectives. Log-likelihood tests show that with the exception of adverbs (LL = 0.10 for 1 d.f., $p = 0.757$), the differences between native and translational Chinese in other content word classes are all significant ($p < 0.001$).

Figure 5 shows the distribution of major function word classes in the two corpora. Log-likelihood tests indicate that with the exceptions of space words (LL = 0.55 for 1 d.f., $p = 0.458$) and non-predicate modifiers (LL = 0.43 for 1 d.f., $p = 0.512$), all other word classes display a highly significant difference ($p < 0.001$). As can be seen from Figure 5, classifiers, modal particles, time words, place words and descriptive words are all more commonly used in native Chinese. Descriptive words are similar to adjectives, while place words and time words are often considered as subsets of nouns, which explains why they display similar patterns of distribution to adjectives and nouns respectively. Given that Chinese is a classifier language whereas English is not, with classifiers used 29 times as frequently in Chinese as in English (see Xiao and McEnery 2010), it is unsurprising to find that classifiers are under-used in ZCTC (see Section 4.4 below), which contains text samples that are mostly translated from English. Native Chinese makes more frequent use of modal particles possibly because they are a unique item in Chinese: modal particles such as 吗 *ma*, 呢 *ne*, 啊 *a* and 吧 *ba* can hardly find formal equivalents in English (see Section 4.7).

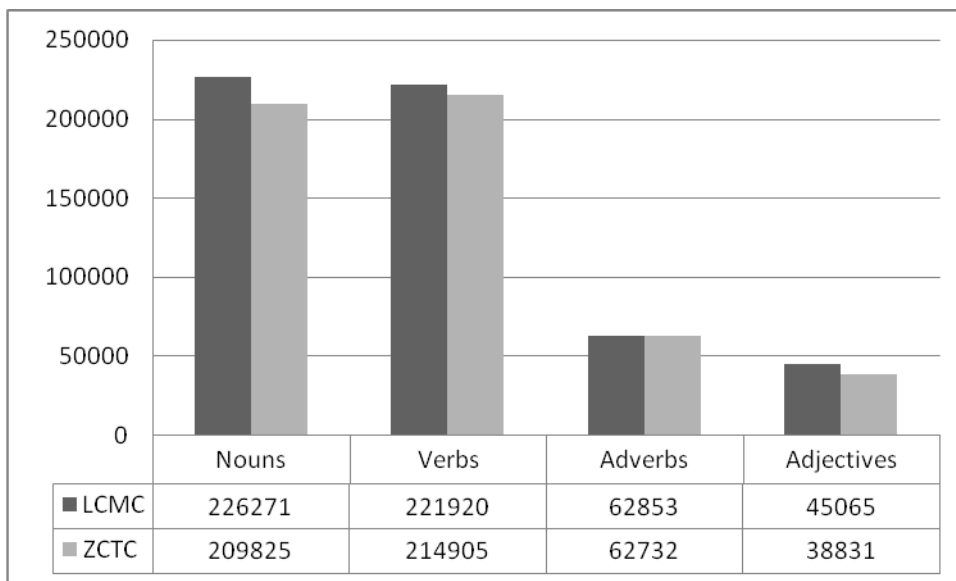


Figure 4. Content word classes in LCMC and ZCTC

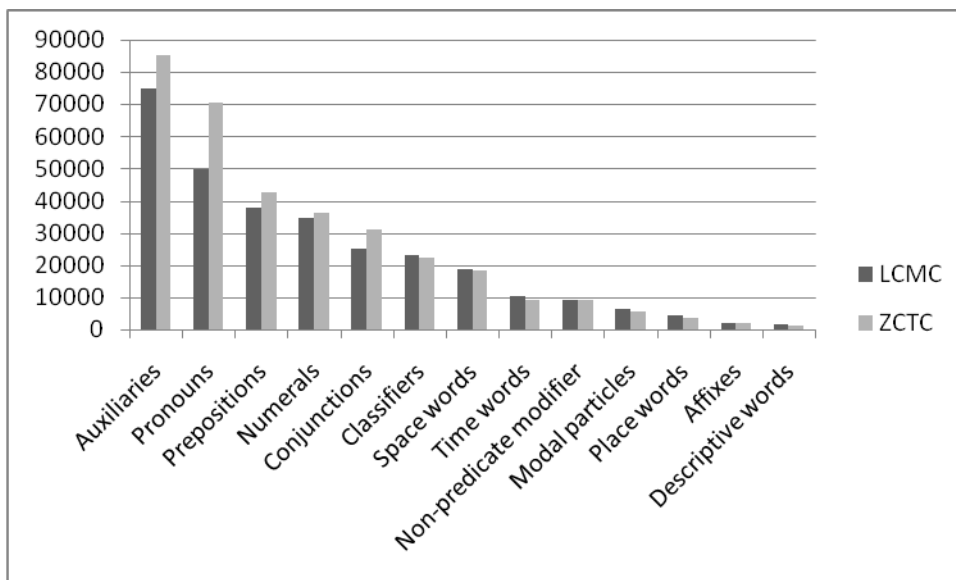


Figure 5. Major function word classes in LCMC and ZCTC

In contrast, auxiliaries, pronouns, prepositions, numerals and conjunctions are all more commonly used in translational Chinese. Given that the frequency of function words is generally higher in translational language, auxiliaries are unsurprisingly more frequently used in translational than native Chinese. Because Chinese is not a morphologically inflectional language, the more frequent use of prefixes and suffixes (e.g. 超 *chao*- ‘super’, 们 *-men* ‘plural suffix’, 者 *-zhe* ‘-er/-or/-ist’, 性 *-xing* ‘-ness/-ity’) in translated Chinese texts is arguably a result of the influence of English source texts.

Xiao (2010) finds that the mean frequency of conjunctions in ZCTC is greater than in LCMC (306.42 vs. 243.23 instances per 10,000 tokens), a difference that is statistically significant (LL = 723.12 for 1 d.f., $p < 0.001$), while translational Chinese is also found to show a preference for high-frequency conjunctions and more colloquial and simpler conjunctions. These findings provide evidence in support of the explicitation and simplification hypotheses. In the section to follow, we will focus on two major function word classes, i.e. pronouns and prepositions.

3.4 Pronouns and prepositions

As pronouns have the function of making discourse more cohesive as noted earlier, translational language is hypothesised to make more frequent use of pronouns. This section will seek to test this hypothesis by exploring the overall distribution of pronouns in LCMC and ZCTC.

As can be seen from Figure 6, which shows their overall distribution, pronouns are distributed in native and translational Chinese in a similar pattern, with the most frequent use in fiction, followed by general prose and news, and the least frequent use in academic prose. However, translational Chinese makes more frequent use of pronouns no matter whether the two corpora are taken as a whole or individual registers are considered. All differences are statistically significant ($p < 0.001$) according to the results of log-likelihood tests.

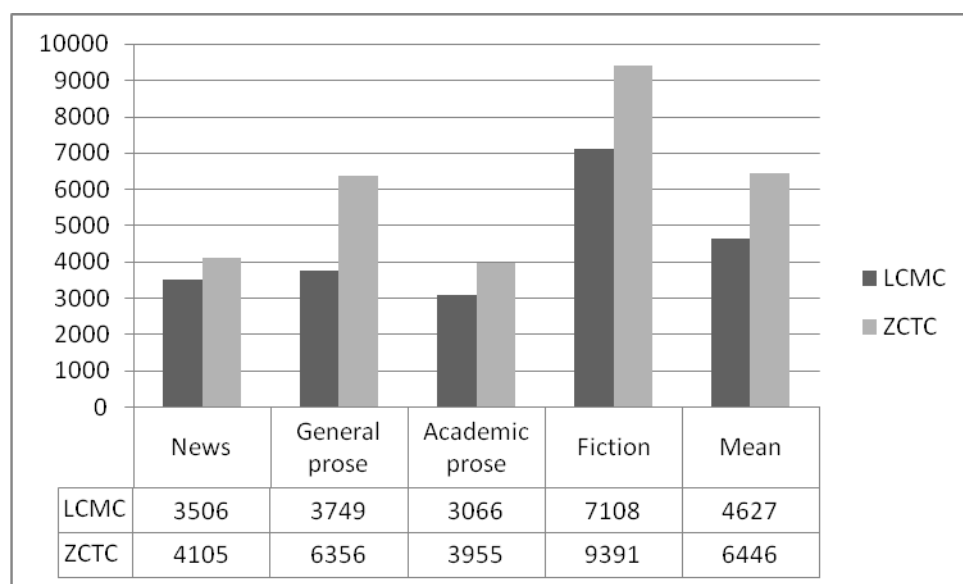


Figure 6. Pronouns in LCMC and ZCTC

The significantly more frequent use of pronouns (especially personal and demonstrative pronouns) can be regarded as an indicator of translational explicitation (cf. Xiao 2012). The relatively lower frequency of pronouns in LCMC than in ZCTC can be accounted for by the fact that in native Chinese, unlike in English, the grammatical subject can be dropped because of its connective discourse function, whereas the subject in the English source text is likely to be transferred to the translated text. This point is well illustrated in example (1a), which is excerpted from *A Mandman's Diary* by the renowned Chinese writer Lu Xun.

(1a) 我看不见他，已经三十年多年了；今天[我]见了，精神分外爽快。[我]才知道以前的三十多年，[我]全是发昏，然而[我]须十分小心。

Wo kanbujian ta, yijing sanshi duo nian le; jintian [wo] jian le, jingshen fenwai shuangkuai. [Wo] cai zhidao yiqian de sanshi duo nian, [wo] quan shi fahun, ran'er [wo] xu shifen xiaoxin.

(1b) I have not seen it for over thirty years, so today when I saw it I felt in unusually high spirits. I begin to realize that during the past thirty odd years I have been in the dark; but now I must be extremely careful.

In example (1a), which is originally written in Chinese, the subject pronoun 我 *wǒ* ‘I’ is dropped after its occurrence in the first sentence. Although the passage comprises more than one sentence, the subject pronoun in the first sentence functions to glue the ensuing discourse in the excerpt together. Because of the cohesive function of pronouns in Chinese, a competent Chinese speaker would hardly have any difficulty in understanding the passage. However, if the same message is translated into Chinese from English (1b), the translator is very likely, under the influence of the English source text, to include all of the dropped subjects as highlighted and included in the brackets in (1a). This is because English and Chinese have different conventions of using pronouns: English tends to repeat personal pronouns, which is dispreferred in Chinese so that where a pronoun is repeated in a text in English, Chinese either drops the pronoun or repeats a noun instead (cf. Liu 1991: 371).

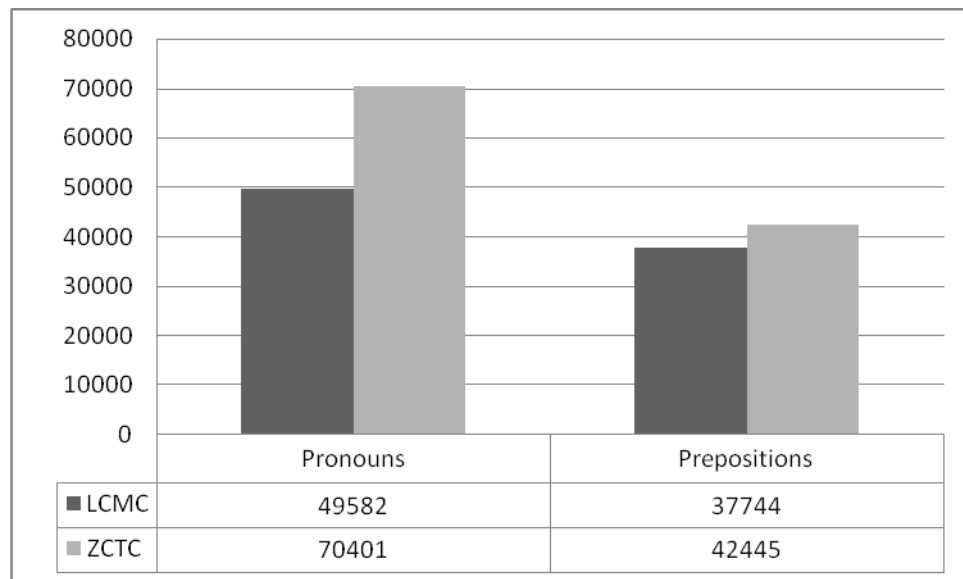


Figure 7. Pronouns and prepositions in LCMC and ZCTC

Prepositions as a major function word class also facilitates explicitation at the grammatical level by explicating grammatical relationships. As can be seen in Figure 7, while the contrast between native and translational Chinese is not as marked as in pronouns, prepositions are also significantly more frequent in translated texts (LL = 293.29 for 1 d.f., $p < 0.001$).

3.5 Idioms

According to Sinclair (1991), the idiom principle operates in combination with the open choice principle to mirror the distinction between conventionality and flexibility in language use. Of the two, the idiom principle plays a central role in speech and writing, relying heavily on the speaker or writer’s large inventory of prefabricated lexico-grammatical chunks at their disposal. Idioms are pre-packaged, recurring stretches of language (Baker 2004). They are fixed or semi-fixed lexical phrases closely associated with idiomaticity and fluency, which, according to the TU hypothesis of normalisation, are a “preferred strategy” that translators tend to adopt (Baker 2004: 182). This section will compare the use of idioms in native and translational Chinese to verify this hypothesis.

Figure 8 shows the distribution of idioms in LCMC and ZCTC. It can be seen that in both native and translational Chinese, idioms have the highest frequency of occurrence in news, followed by general prose and fiction, and have the lowest frequency in academic prose. It is also clear that no matter whether the two corpora are taken as a whole or the four broad registers are examined separately,

idioms are more frequently used in native than translational Chinese. Log-likelihood tests indicate that both the overall difference and the differences in individual registers are highly significant, with $p = 0.006$ for fiction and $p < 0.001$ for all other registers and the overall difference. The substantially more common use of idioms in native than translational Chinese suggests that the TU hypothesis of normalisation is unsupported by our corpus data.

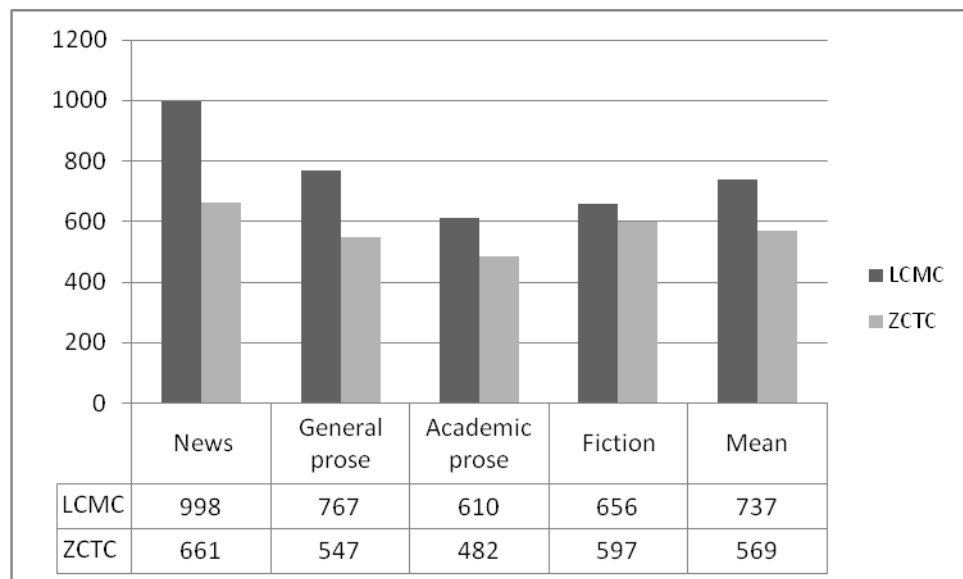


Figure 8. Idioms in LCMC and ZCTC

3.6 Punctuation

Punctuation can be useful in investigating the properties of translational language. This section will compare the use of major types of punctuation in LCMC and ZCTC. There are two types of punctuation, namely sentence-final (full stop, question mark, semi-colon, and exclamation) and non-sentence-final (all other punctuation marks). However, not all the above kinds of punctuation are useful in this research. For example, the frequency of quotation marks and elliptical marks merely indicates the amount of quotations and ellipsis while the Chinese-style centred dot mark is unavoidable, instead of as an optional style marker, in translating foreign personal names (as in 保罗·德雷克, the Chinese translation of *Paul Drake*). Hence these punctuation marks will not be investigated in this section. A colon is mainly used to introduce a quotation, though it can also function as an explicating device when it is used to introduce a list of items. But because these two different uses are not separately annotated in our corpora, colons will not be included in our analysis either.

Figure 9 shows the normalised frequencies (per 100,000 words) of major kinds of punctuation in native and translational Chinese. Log-likelihood tests show that the differences in the frequencies of all of these punctuation marks are statistically significant (with $p = 0.026$ for question marks and $p < 0.001$ for all other punctuation marks). As can be seen, of the four sentence-final punctuation marks, exclamations, question marks and semi-colons are more frequent in native Chinese whereas full stops are more frequent in translational Chinese.

This is because in native Chinese texts, complete sentences do not always end with full stops, because commas are often used to replace full stops whereas full stops in English source texts tend to be transferred into translated Chinese texts, which explains why full stops are less frequent but commas are more common in LCMC. Of the non-sentence-final punctuation marks, the pause mark in Chinese is used to indicate a pause between words and phrases within a sentence. The pause mark is non-

existent in English, which uses the comma where a pause mark is used in Chinese. In English-to-Chinese translation, a translator is very likely to use the comma by following the English source text rather than use a pause mark. Consequently, the frequency of pause marks in LCMC doubles that in ZCTC. In this regard, it can hardly be said that the normalisation hypothesis is tenable in Chinese. The other two major punctuation marks – parentheses and dashes, can both be used to provide explanations or illustrations (as exemplified by the use of dashes and parentheses in this sentence). Like reformulation markers (see Xiao 2011), they are explicating devices that facilitate semantic explicitation in translation. As such, these types of punctuation can reasonably be expected to occur more frequently in translational Chinese.

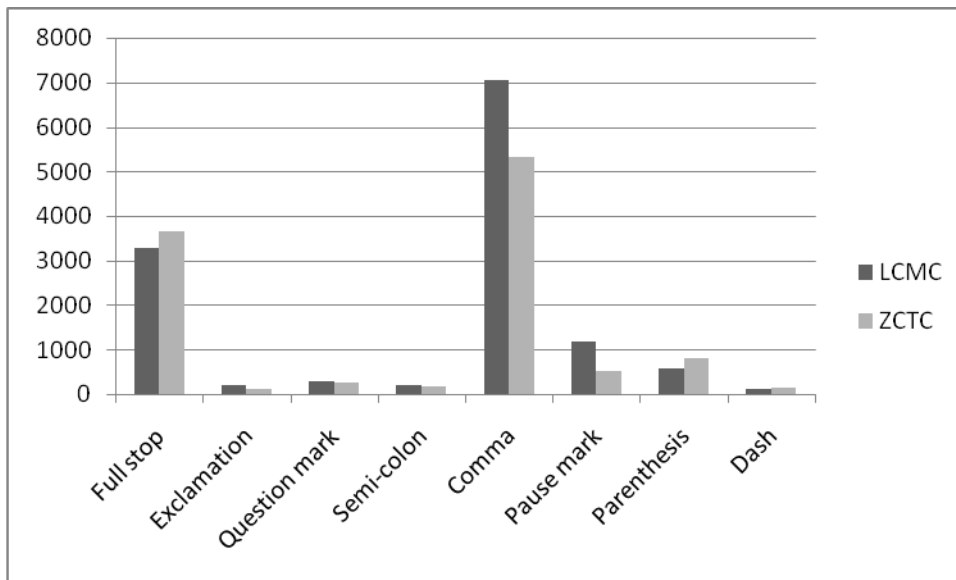


Figure 9. Major punctuation marks in LCMC and ZCTC

4. Grammatical properties of translational Chinese

Having explored the lexical properties of translational Chinese, we will now shift our focus of discussion to the grammatical level by investigating a range of grammatical properties in native and translational Chinese, including mean sentence segment length (Section 4.1), basic sentence structures such as passives marked by 被 *bei* and 为...所 *wei...suo* (Section 4.2) and disposal 把 *ba* constructions (Section 4.3), as well as grammatical categories such as classifiers (Section 4.4), aspect markers (Section 4.5), structural auxiliaries (Section 4.6), and modal particles (Section 4.7).

4.1 Mean sentence segment length

Mean sentence length has often been used as a parameter in research of translational language. However, different results have been reported in different studies. For example, Malmkjær (1997) observes that using stronger punctuations in translation entails shorter sentences in translational language, while Laviosa (1998b) notes that, in comparison with comparable native English, the mean sentence length is lower in translated English news texts but higher in translated English literary texts. According to Xiao (2010), while the mean sentence length is slightly greater in LCMC than in ZCTC, the difference has no statistical significance ($t = -1.41$ for 28 d.f., $p = 0.17$). Mean sentence length also allows different, and even conflicting, interpretations. As Pym (2008) argues, shorter sentences in translated texts are expected according to the simplification hypothesis whereas explicitation calls for

longer sentences; it is even possible to argue that a sentence that becomes longer because of explicitation is actually simpler to read so that both longer and shorter sentences can be taken, contradictorily, as an indicator of simplification.

As noted in Section 3.6, Chinese sentences do not always end with a full stop. Chen (1994) finds that three quarters of sentences ending with a full stop and semi-colon contain two or more complete sentence segments. For example, in example (2),⁴ while the Chinese version only contains one sentence, it actually expresses three relatively complete meanings, and as such, three sentences are used in the English version.

(2a) 人们大多通过电影认识叶锦添, (2)尤其在 2001 年凭着《卧虎藏龙》中典雅清幽的东方意象, 夺下华人世界第一座奥斯卡最佳美术指导奖后, (3)各地蜂拥而至的邀约, 更快速将他推向全球舞台。

Renmen daduo tongguo dianying renshi Ye Jintian, (2) youqi zai 2001 nian ping-zhe “Cang Long Wo Hu” zhong dianya qingyou de dongfang yixiang, duoxia huaren shijie diyi zuo Aosika zui jia meishu zhidao jiang hou, (3) gedi fengyong'erzhi de yaoyue, geng kuaisu jiang ta tuixiang quanqiu wutai.

(2b) Most people know Tim Yip through films. (2) In particular, in 2001 he became the first ever person from the Chinese world to win the US Academy Award for Best Art Direction, received for the elegant Oriental imagery he brought to Crouching Tiger, Hidden Dragon. (3) Since then, demand for his services has gone into hyperdrive, accelerating the spread of his fame and appeal worldwide.

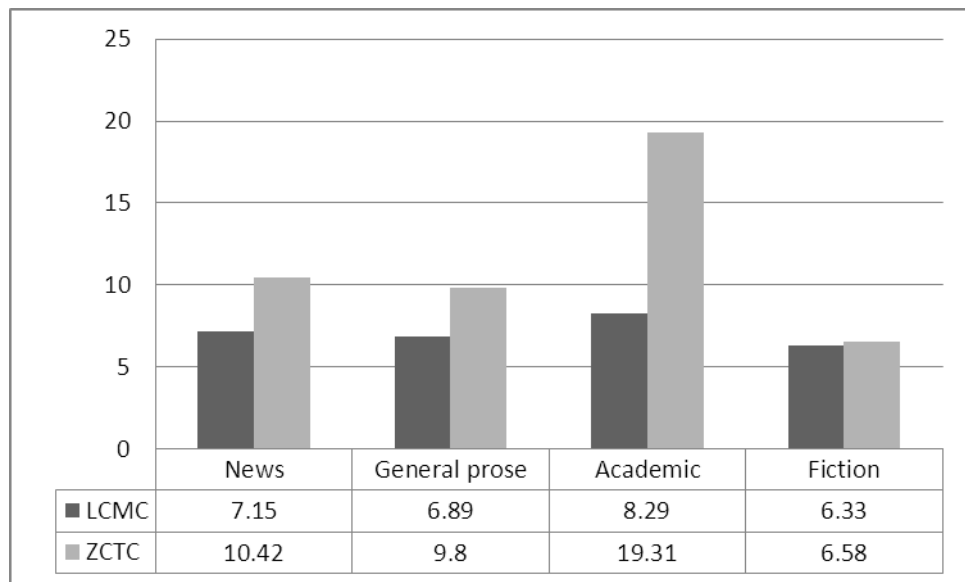


Figure 10. Mean sentence segment lengths in LCMC and ZCTC

Hence, Wang and Qin (2010: 169) argue that for languages that are characterised by parataxis such as Chinese (Liu 1991), sentence segment length is more meaningful than sentence length. This section will compare the mean sentence segment lengths in native and translational Chinese.⁵ Figure 10 compares the mean sentence segment lengths of native and translated Chinese texts. Clearly, the mean sentence segment length is greater in translational Chinese than in native Chinese, in all of the four

registers., with the most marked contrast in academic prose because the corresponding register in English customarily makes use of long sentences. This finding is in line with Wang and Qin's (2010: 169) observations of literary and non-literary translations. One possible explanation is source language interference, because the mean sentence segment length in English is greater than that in Chinese (the mean sentence segment length is 25.59 words in FLOB but only 13 words in LCMC).

4.2 Passives with 被 *bei* and 为...所 *wei...suo*

Passives in Chinese can be syntactically marked with 被 *bei*, 叫 *jiao*, 让 *rang*, 给 *gei* and the archaic structure 为...所 *wei...suo*, but *jiao*, *rang* and *gei* are not fully fledged passive markers (cf. Xiao et al. 2006). This section only considers passives marked with *bei* and *wei...suo*. Passives that profile the agent are conventionally called 'long passives' while those that do not are known as 'short passives'. *Bei* passives can take either long or short form whereas *wei...suo* can only be used in the long form.

Figure 11 shows the proportions of short and long passives in native and translational Chinese, and for comparative purposes, the corresponding figures in the native English corpus FLOB are also included. As can be seen, although short passives are more frequent than long passives in both native and translational Chinese, the proportion of short passives in ZCTC is significantly greater than in LCMC (LL = 63.1 for 1 d.f., $p < 0.001$). The higher proportion of short passives in translational Chinese is clearly a result of source language interference, because the short passive is the statistical norm of passive use in English (see Xiao et al. 2006), which accounts for over 90% of the total, as shown in Figure 11. The passive in English is a strategy for expression in that it is used when the agent is unknown or there is no need to mention the agent. In Chinese, in contrast, three out of the five syntactic passive markers (*wei...suo*, *jiao*, *rang*) can only occur in long passives, while the proportions of short passives for the other two (60.6% and 57.5% for *bei* and *gei* respectively) are considerably lower than that of English passives (cf. Xiao et al. 2006). As earlier Chinese grammarians Lü and Zhu (1979) and Wang (1985) noted, the agent must be included in the Chinese passive, though this constraint has become more relaxed. When it is hard to identify the agent, vague expressions such as *ren* 'person, someone' or *renmen* 'people' is specified as the agent, which seldom occurs in English passive use. In cases where English uses the passive but does not profile the agent, Chinese tends to avoid the passive.

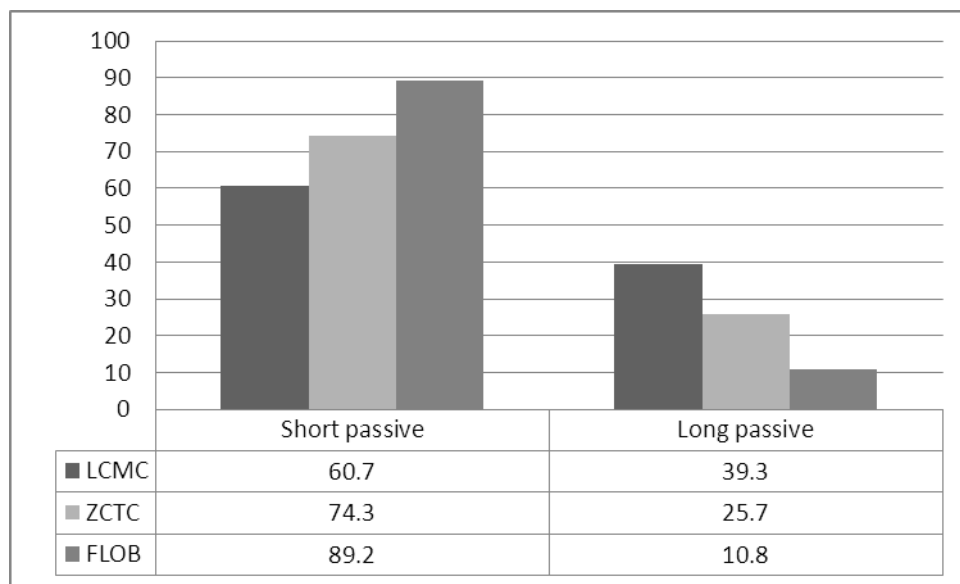


Figure 11. Short and long forms of *bei* passives

Figure 12 compares the pragmatic meanings expressed by *bei* passives in LCMC and ZCTC and by English *be* passives in FLOB. As can be seen, there are significant differences in the proportions of different pragmatic meanings between the three corpora (LL = 212.28 for 2 d.f., $p < 0.001$), with the translational Chinese corpus positioned between the native Chinese and native English corpora, and particularly marked contrasts in neutral and negative meaning categories. Passives in English and Chinese have different functions. English passives primarily function to mark a formal, objective and impersonal style, and are thus pragmatically neutral whereas Chinese passives are an “inflictive voice” that tends to express a negative pragmatic meaning, evaluating the event being described as undesirable, unfavourable or adversative (Xiao et al. 2006). This is because the prototypical passive marker 被 *bei* is derived from a verb in ancient Chinese which meant ‘suffer’. Consequently, many disyllabic words with 被 *bei* in modern Chinese refer to something undesirable, e.g. 被捕 *beibu* ‘be arrested’, 被俘 *beifu* ‘be captured’, 被告 *beigao* ‘the accused’, 被害 *beihai* ‘be victimised’, and 被迫 *beipo* ‘be forced’, though the semantic constraint on passive use in modern Chinese is no longer as rigid as before (Xiao et al. 2006).

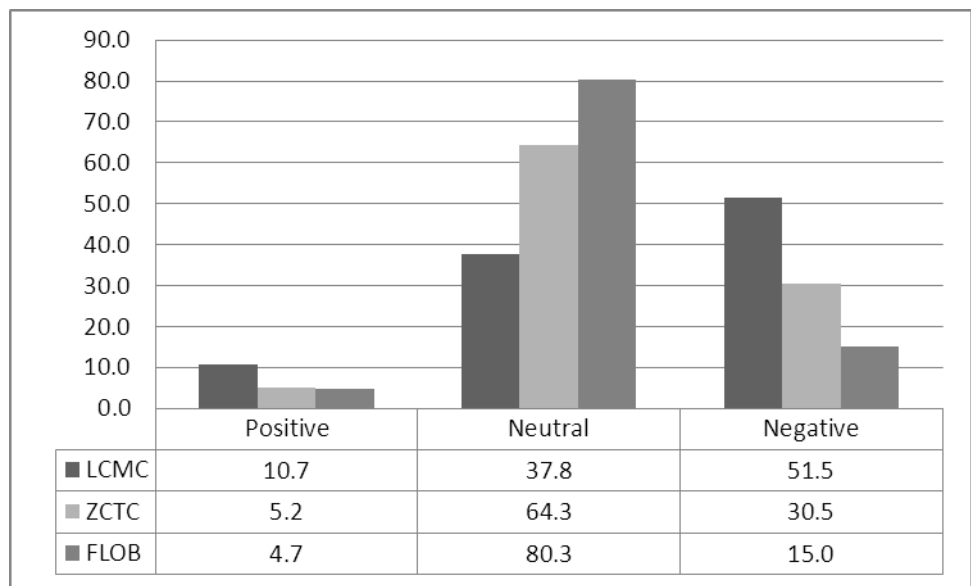


Figure 12. Pragmatic meanings expressed by *bei* passives

Native and translational Chinese also differ in the overall frequency of passive use and in the distribution of passives across genres. Figure 13 shows the normalised frequencies of passives in different genres in the two Chinese corpora. It is clear that the overall mean frequency of passives is significantly greater in translational Chinese than in native Chinese (LL = 69.59 for 1 d.f., $p < 0.001$). Given that passives are over ten times as frequent in English as in Chinese (Xiao et al. 2006: 141-142), it is hardly surprising that translated Chinese texts in ZCTC (99% translated from English) make more frequent use of passives than original Chinese writings. It can also be seen that the most marked contrasts between native and translational Chinese in the distribution of passives are in the genres of reports and official documents (H), news reviews (C) and academic prose (J), where passives are significantly more frequent in translational Chinese, and in detective stories (L), where passives are substantially more frequent in native Chinese. This is because the first three genres mentioned are all formal writings in English which tend to overuse passives as a style marker, which may be transferred into translated Chinese texts, while detective stories are largely concerned with victims who suffer from various kinds of inflictive events, as described by Chinese passives.

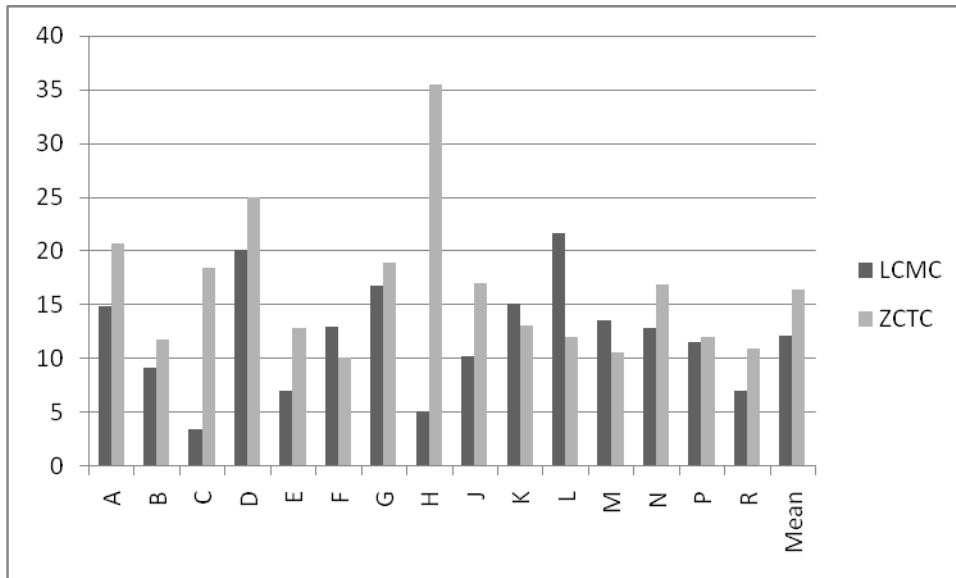


Figure 13. Distribution of *bei* passives in LCMC and ZCTC

The differences between native and translational Chinese in their use of *bei* passives as discussed above can reasonably be regarded as the result of source language interference arising from cross-linguistic differences between English and Chinese (Dai and Xiao 2011). In contrast to *bei* passives, the archaic passive form *wei...suo* is significantly more frequent in native Chinese (with 74 and 50 instances in LCMC and ZCTC respectively; $LL = 5.14$, for 1 d.f., $p = 0.023$), which leads to the speculation that translational Chinese tends to avoid archaic forms in favour of simpler modern forms (see Section 4.6 for further discussion).

4.3 Ba constructions

The disposal 把 *ba* is one of the most important and commonly used constructions in Chinese. It is a unique sentence structure in Chinese, which can hardly find any equivalent in European languages such as English. Only verbs with a proposal meaning can occur in the *ba* construction, where *ba* is a preposition that moves the grammatical object of a verb from its normal position following the verb to a pre-verbal position to highlight the object as well as the action denoted by the verb and its result. The disposal construction also has the cohesive discourse function as moving the verbal object away from its usual position frees the verb up so that it can combine with other sentential elements more closely to express complex ideas. This section compares the use of the *ba* construction in native and translational Chinese.

Figure 14 shows the distribution of *ba* constructions in terms of normalised frequency (per 100,000 words). It can be seen that the overall mean frequencies in the two corpora are close to each other, a difference without statistical significance ($LL = 0.87$ for 1 d.f., $p = 0.351$), but there are register variations between native and translated texts. *Ba* constructions are most frequent in fiction and least frequent in academic prose in both native and translational Chinese because *ba* sentences are descriptive in nature. In fiction, *ba* constructions are more common in translational Chinese whereas in the other three registers, they are more frequent in native Chinese, though the contrast in general prose (the difference is not significant) is not as marked as in news and academic prose. As the *ba* construction is a colloquial feature, and has the textual function of cohesion at the discourse level, it can be used to make text easier to read (cf. Xiao 2012). The more frequent use of *ba* constructions in

translated fiction, which is for light reading, can be considered as an indicator of simplification. When a register is not for light reading, e.g. in news and academic writing, *ba* constructions are under-represented in translation.

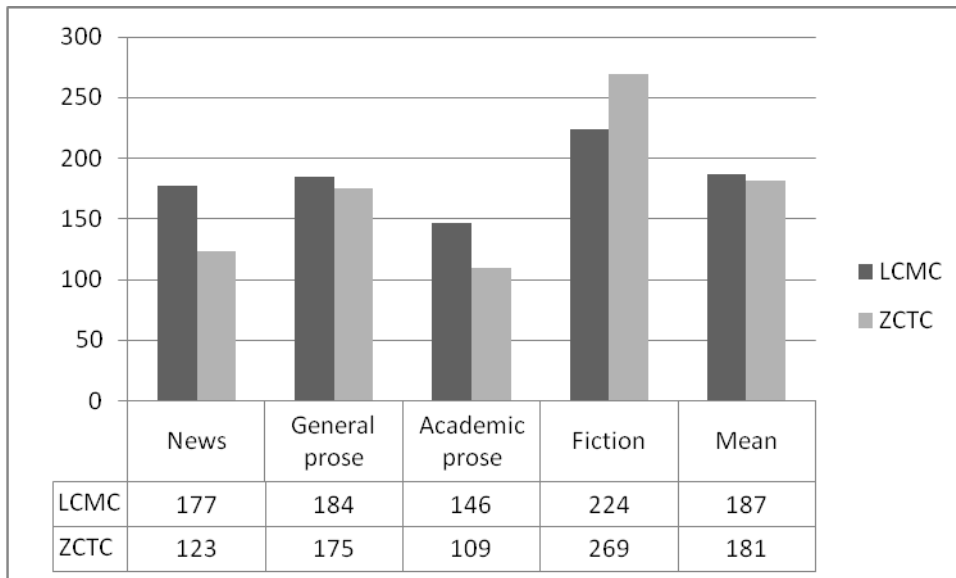


Figure 14. *Ba* constructions in LCMC and ZCTC

4.4 Classifiers

Chinese is well accepted as a classifier language in which the use of classifiers is mandatory. It has a very well developed classifier system that comprises three broad categories: nominal (q), verbal (qv) and temporal (qt), which are used respectively to quantify nominal entities, verbal actions and time (Xiao and McEnery 2010).

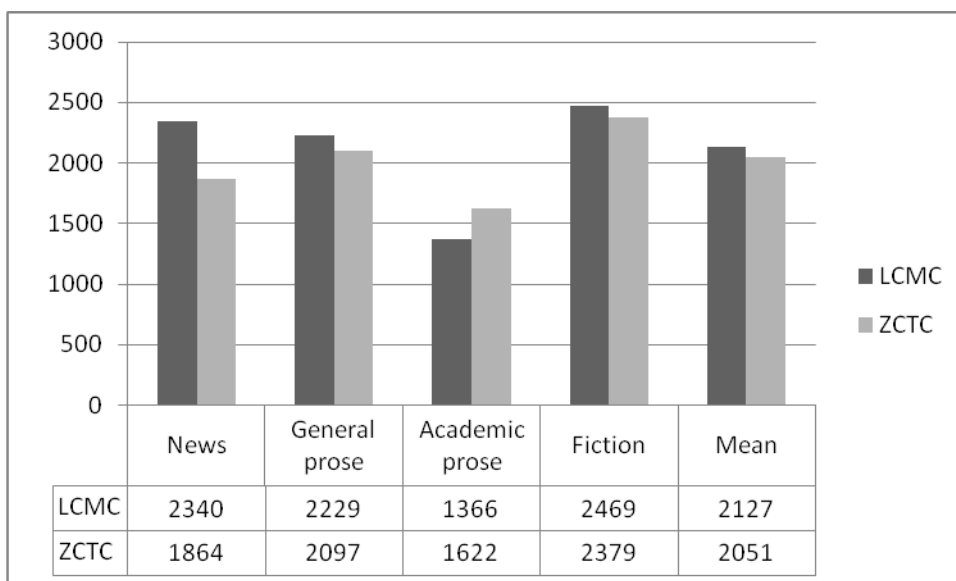


Figure 15. Classifiers in LCMC and ZCTC

Figure 15 shows the normalised frequencies (per 100,000 words) of classifiers in native and translational Chinese, including their distribution in the four broad registers. It is clear that the overall

frequency of classifiers is higher in native than translational Chinese (LL = 36.66 for 1 d.f., $p < 0.001$). Classifiers are also significantly more frequently used in native Chinese in all registers other than academic prose, reflecting the cross-linguistic difference in the status of classifiers in Chinese and English: English is not a classifier language, in which the use of classifiers is only required for noncount nouns, which explains why classifiers are 29 times as frequent in Chinese as in English (Xiao and McEnergy 2010). In the register of academic prose, in contrast, classifiers are significantly more frequent in translational Chinese. A cross-tabulation of registers and classifier categories can help to account for this different distribution pattern.

As shown in Figure 16, in the news register all three types of classifiers are more frequent in native Chinese, and log-likelihood tests indicate that these differences are statistically significant ($p \leq 0.001$). In general prose, all types of classifiers are also more frequent in native Chinese, but only the difference in verbal classifiers (qv) is significant ($p < 0.001$). In academic prose and fiction, only the difference in temporal classifiers (qt) is significant ($p < 0.001$), with fiction displaying a more frequent use of classifiers in native Chinese and academic prose showing a more frequent use of classifiers in translational Chinese. The cross-tabulation in Figure 16 suggests that academic prose alone makes significantly more frequent use of one type of classifiers, namely temporal classifiers (qt). Concordances show that the temporal classifier that makes the difference is 年 *nian* ‘year’, which is mainly used in citing statistics and making references in academic writing, e.g. 根据 1997 年 6 月 BIS 的报告 ‘According to the BIS report in June 1997’, and 正像 1991 年的《世界发展报告》所指出的那样 ‘As pointed out in *World Bank* (1991)’. This classifier alone takes up 80% of temporal classifiers in academic prose in the translational corpus, which doubles the corresponding percentage in the same register in the native Chinese corpus. Clearly, the significantly more frequent use of this temporal classifier in academic prose in translated Chinese texts can be attributable to the more rigorous academic referencing and citation practice in English source texts. Because of this cultural difference, the use of classifiers in academic prose might be considered as a special case, which will not invalidate the finding that classifiers are generally more common in native Chinese. In other words, translational Chinese is characterised by under-representation of classifiers.

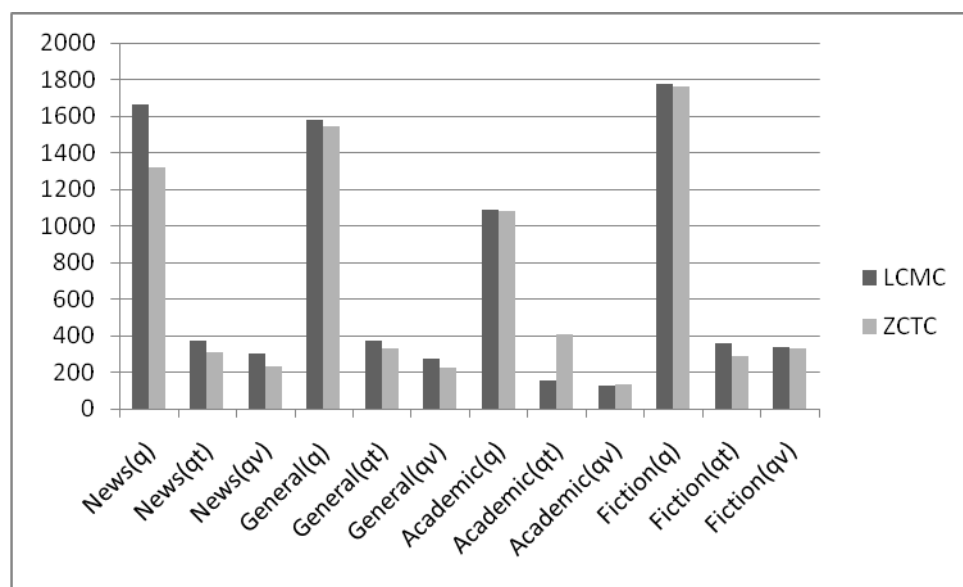


Figure. 16. Genre variations in the distribution of classifiers

4.5 Aspect markers

Chinese as an aspect language relies heavily on aspect markers to express temporal and aspectual meanings. This section investigates the use of three well established aspect markers, namely perfective markers 了 *-le* and 过 *-guo* and the imperfective aspect marker 着 *-zhe*, in native and translational Chinese.

As shown in Figure 17, which illustrates the distribution of the three aspect markers in the two Chinese corpora, *-le* and *-zhe* are significantly more frequent in native Chinese ($p < 0.001$) though *-guo* does not show a significant difference ($p = 0.752$), possibly because of its low overall frequency of use in the two corpora. Hence it can be concluded that, in general, aspect markers are under-represented in translational Chinese.

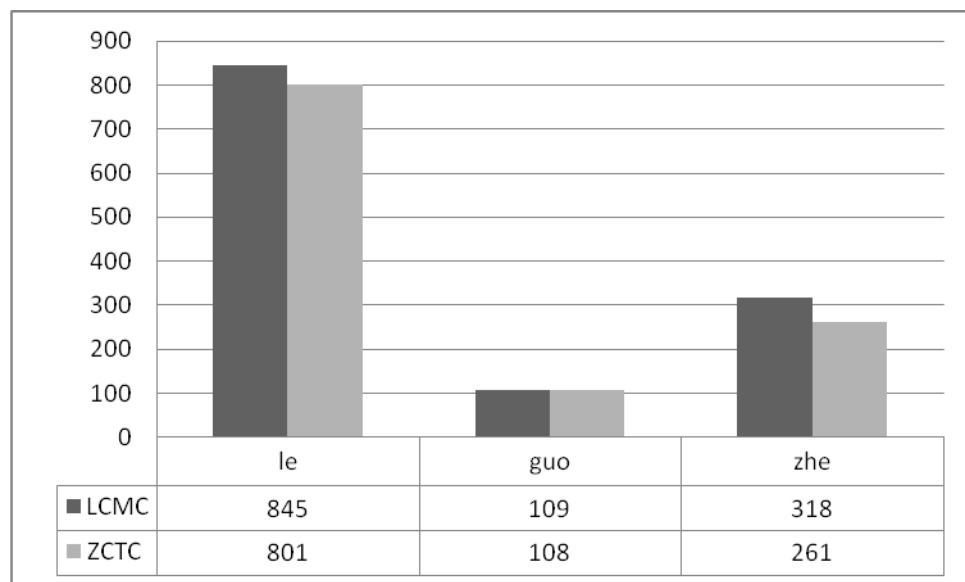


Figure 17. Aspect markers in LCMC and ZTC

4.6 Structural auxiliaries

This section discusses the distribution of structural auxiliaries 的 *de*, 地 *de*, 得 *de*, and 之 *zhi*. These function words have no content meaning but are used to join two words, or are used to follow a particular word to indicate a certain grammatical structure or relational meaning. More specifically, 的 *de* is the attributive marker that joins two words to form an adjectival endocentric structure functioning as an attributive modifier.⁶ 地 *de* is the adverbial marker that joins two words to form an adverbial or verbal endocentric structure functioning as an adverbial modifier. 得 *de* is the complemental marker that joins two words to form an adjectival or verbal endocentric structure functioning as a complement following a verb. 之 *zhi*, like 的 *de*, is also an attributive marker, which is an archaic structural auxiliary handed down from ancient Chinese and used in modern Chinese to form the four-character structures or to replace 的 *de* to avoid repetition.⁷

Figure 18 compares the normalised frequencies (per 100,000 words) of the three modern structural auxiliaries in native and translational Chinese. As can be seen, no matter whether the two corpora are taken as a whole or literary and non-literary components are considered separately, the three modern structural auxiliaries are significantly more frequent in translated texts ($p < 0.001$). The result may be unexpected given that these words are unique in Chinese and have no formal equivalents

in English.⁸ However, because these structural auxiliaries are all highly frequent function words in Chinese, it is words such as these that function to facilitate structural explication. In this sense, it is a quite natural expectation for these high-frequency function words to occur more frequently in translational Chinese.

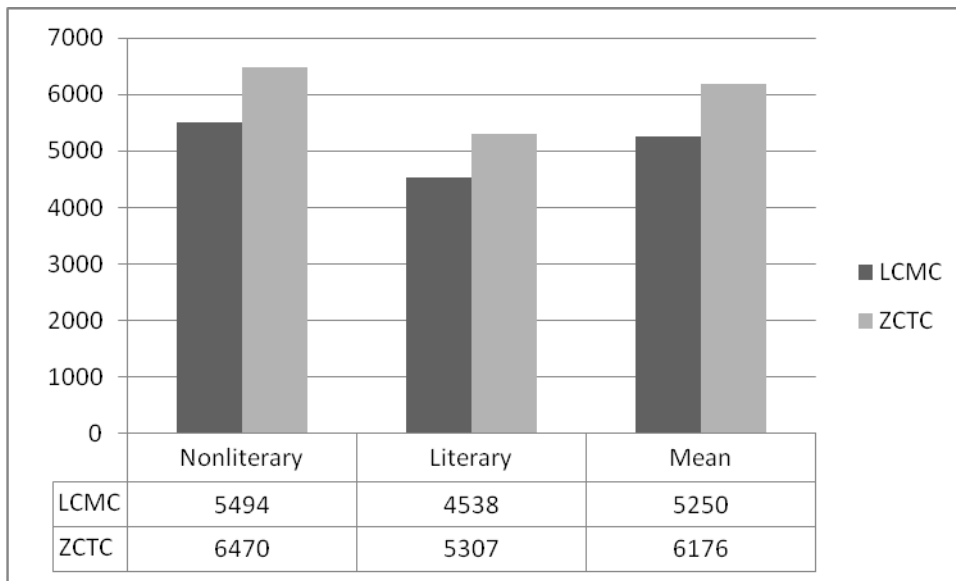


Figure 18. Structural auxiliaries 的 *de*, 地 *de* and 得 *de* in LCMC and ZCTC

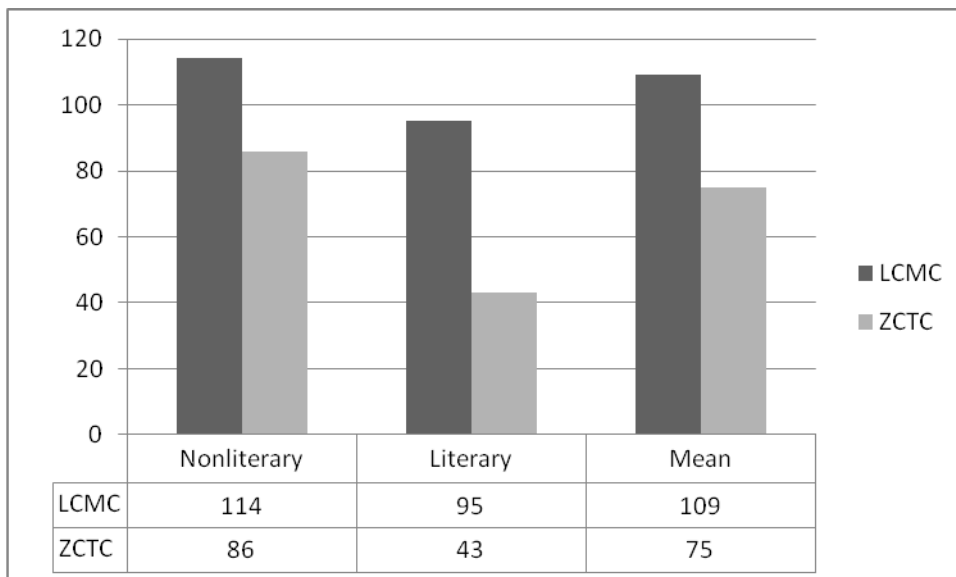


Figure 19. Archaic structural auxiliary 之 *zhi* in LCMC and ZCTC

The archaic structural auxiliary 之 *zhi*, in contrast, displays a totally different pattern of distribution as shown in Figure 19, which shows its normalised frequencies in the two Chinese corpora. While the frequency of 之 *zhi* is not particularly high, the contrast between native and translational Chinese is marked, with a much more frequent use in native Chinese, both in literary and non-literary texts. The significantly less frequent use of 之 *zhi* ($p < 0.001$) appears to be related to its archaic style: translators tend to consciously avoid archaic words and structures in favour of simpler modern forms (see also Section 4.2), which might be taken as a manifestation of translational simplification.

4.7 Modal particles

This section discusses another unique feature of Chinese, namely modal particles, which are used at the end of a sentence to express the speaker's mood or attitude. Commonly used modal particles in Chinese include, among other, 吗 *ma*, 呢 *ne*, and 吧 *ba*. They have no formal equivalents in English, which uses auxiliaries, modal verbs as well as word order and tones to achieve similar effects.

Figure 20 illustrates the distribution of modal particles in native and translational Chinese. The normalised frequencies (per 100,000 words) show that modal particles are significantly more frequent in native Chinese ($p < 0.001$), in both literary and non-literary texts, suggesting that translational Chinese is indeed different from native Chinese. The less frequent use of modal particles in translational Chinese can possibly be explained by the fact that modal particles are a unique grammatical category in Chinese that is not used in the English source language, thus affecting the use of modal particles in translated texts. In this sense, TL unique item under-representation might as well be regarded as an indicator of source language interference.

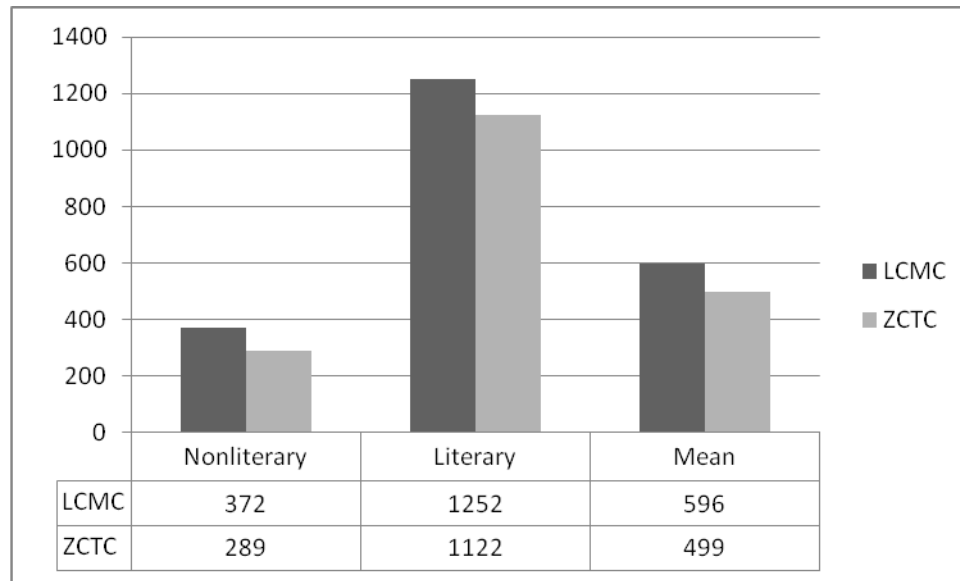


Figure 20. Modal particles in LCMC and ZCTC

5. TU hypotheses reevaluated from the Chinese perspective

This section will reevaluate the TU hypotheses reviewed in Section 2 in the face of evidence arising from translational Chinese, on the basis of the discussions of the lexical and grammatical properties of translational Chinese explored in Sections 3 and 4.

As noted in Section 2, explicitation has been investigated extensively and has been found to be one of the least controversial TU hypotheses and is supported by fresh evidence from translational Chinese in the present study. Explicitation in English-to-Chinese translation is manifested in three aspects: semantic explicitation, grammatical explicitation, and logical explicitation. Semantic explicitation means that meaning is expressed more explicitly in translational Chinese by making more frequent use of reformulation marks (Xiao 2011) and explicating and explanatory punctuation marks such as dashes and parentheses. Linguistic properties that provide evidence in support of grammatical explicitation include more frequent use of function words such as pronouns, prepositions and structural auxiliaries in translational Chinese, which all function to explicate grammatical and structural

relationships, thus resulting in longer sentence segments. In addition, logical explicitation at the discourse level is evidenced by a higher frequency of conjunctions in translational Chinese, which helps to explicate logical relationships between clauses and make translated texts more cohesive. All of these lexical and grammatical properties of translational Chinese suggest that the English-based hypothesis of translational explicitation also holds in translational Chinese. Of course, explicitation and implicitation in translation are relative phenomena as the extent of explicitation or implication is not only affected by linguistic factors but also depends on individual translators' preferences as well as socio-cultural factors (Wang 2005).

Translational Chinese also has many properties that provide evidence supporting the simplification hypothesis. For example, a lower percentage of content words in translational Chinese entails a lower information load, which is an indicator of semantic simplification. The tendency in translational Chinese to repeat high-frequency words and colloquial conjunctions (Xiao 2010) and oral reformulation marks (Xiao 2011) and to avoid using near synonyms alternatively can be taken as a manifestation of lexical simplification. The less frequent use of syntactically complex and archaic structures in translational Chinese is an indication of grammatical simplification. The more frequent use of word clusters of high-frequency and high-coverage in translated texts (see Xiao 2011), together with the various explicitation strategies as noted above, results in simplification at the discourse level. Nevertheless, translational simplification is actually not such a simple phenomenon. Translational Chinese can be said to be a mixture of simplification and complication, meaning that while some linguistic properties of translations are simpler than comparable original Chinese writings, others may make translated texts even more complicated or hard to read. For example, while translational Chinese tends to repeat ultra-high-frequency words, it also makes less frequent use of sub-high-frequency words in terms of word types but makes more frequent use of hapax legomena and other words of extremely low frequency. The mean word length is generally greater in translated texts than in native Chinese texts, which is unavoidable given that transliterations, which are usually longer than native Chinese words, are abundant in translation. In addition, although no significant difference is found in the mean sentence lengths of native and translational Chinese, the latter has longer sentence segments while as noted earlier, mean sentence segment length is more meaningful than mean sentence length for Chinese as a language characterised by parataxis. Furthermore, how to interpret sentence segment length is another issue. As noted in Section 2, mean sentence length can be interpreted differently. A shorter sentence may mean a less complex sentence structure, but it can also mean a more compact and less explicit structure that is even more difficult to understand; conversely, a longer sentence is likely to be a syntactically complicated sentence that is not simple to read, but it is also possible that the sentence is longer because additional words have been used in the translation to make explicit what is implicit in the source text, and thus making the translated message easier to understand. The same can be said of mean sentence segment length in translational Chinese. Hence, it can be concluded that translational simplification is not a pure, simple phenomenon. Some aspects of translational language can be simpler than those in native language, while other aspects may be even more complicated than those in native language. Hence, the TU hypothesis of simplification should take account of the overall synthetic result of the dynamic interplay between simplification and complication.

As Chinese and English are genetically distant languages that have many striking cross-linguistic differences (Xiao and McEnery 2010), unique linguistic features in Chinese can help to differentiate translational Chinese from native Chinese in English-to-Chinese translation, if the TU hypothesis of TL unique item under-representation is tenable. As noted in Sections 3 and 4, unique items in Chinese such as Chinese idioms, pause marks, disposal *ba* constructions, aspect markers, modal particles all demonstrate a tendency of under-representation in translational Chinese, despite register variations for some of the features. In addition, while the structural auxiliary 的 *de* as a highly frequent function word is more common in translational Chinese, the archaic form with a similar

function, 之 *zhi*, is under-represented in translated texts. Likewise, while the *bei* passive is more frequently used in translation, possibly because of source language interference, the archaic passive form *wei...suo* is under-represented in translational Chinese. It is clear that as far as translated texts in English-to-Chinese translation are concerned, the more characteristic a linguistic feature is of Chinese, the better it can serve to differentiate between native and translational Chinese. It appears that the hypothesis of TL unique item under-representation has not received adequate attention in TU research. This limitation in the status quo seems to be due to the fact that TU research has largely been based on and confined to closely related European languages, in which some linguistic features may not be as markedly dissimilar as in genetically distinct languages such as English and Chinese.

While translating can certainly convey messages between languages, given structural differences between languages and underlying cultural differences, a translated text that uses the target language to describe what takes place in the SL culture can hardly achieve the same level of naturalness as a non-translated text that uses the native language to describe an event in its own culture. This unnaturalness is translationese, which is attributable to SL interference and TL unique item under-representation as noted above. Translational language can be said to be a mixture of these two translation phenomena (cf. Teich 2003). In English-to-Chinese translation, SL interference has been observed in a range of linguistic features. For example, translational Chinese follows the SL norm of using full stops instead of using commas to break up sentences into sentence segments as in native Chinese, thus resulting in greater mean sentence segment length. As transliterations are inevitable in translation, the mean word length is generally greater in translational Chinese. Similarly, the preference for using high-frequency, high-coverage word clusters in translational Chinese is also a result of SL interference. Lexical properties such as the more frequent use of prefixes and suffixes, pronouns and light verbs also reflect the influence of the English source language. At the grammatical level, the whole range of differences demonstrated by the *bei* passive between native and translational Chinese, as discussed in Section 4.2, are all manifestations of SL interference.

Normalisation is a highly debatable TU hypothesis. Since it was put forward by Baker (1996), a number of translational features have been used to espouse this hypothesis. Baker (2007) herself uses idioms to illustrate her point, but unfortunately she observes two conflicting tendencies in using idioms in English translations. On the one hand, idioms are supposed to be used heavily in translation to conform to the norm of the target language while on the other hand idioms, especially those characterised with a high degree of opacity, are expected to be avoided in translation because of their informal tone. Evidence arising from the use of Chinese idioms does not support normalisation either (Xiao and Dai 2010). The distribution of pause marks, commas, and sentence-final punctuation marks in translational Chinese also invalidates the claim of using stronger punctuation marks in translation to replace weaker ones in source texts as evidence in support of normalisation. In addition, the preference in translation to repeat high-frequency word clusters or multiword units arguably serves better as evidence in support of simplification than normalisation. More importantly, the discussions of TL unique item under-representation and SL interference show that the differences between native and translational languages are systematic rather than random and occasional. Such systematic differences render translational language a third code that is different from both source and target languages. Even though in literary translation, the dialects spoken by different characters in the source texts may be translated into standard target language (Section 2.3), this limited evidence, which might as well be viewed as a means of simplification, is inadequate for such a strong claim that translational language is characterised by “a tendency to exaggerate features of the target language and to conform to its typical patterns” (Baker 1993: 183), because a wide range of distinctions between native and translational languages cannot be explained away easily by the normalisation hypothesis.

In sum, translated texts share a number of common properties. For example, they are more explicit in meaning, grammatical structure and logical relationship. Translators may try various means

to render them simpler and easier to read. Given TL unique item under-representation and SL interference, translations from the same or similar source languages may share even more common features, which make translated texts more similar to each other than to native texts in the target language, thus lending evidence to the levelling out hypothesis. Nevertheless, levelling out or convergence depends to a great extent on the typological distance between the languages involved in translation, and it can involve many linguistic features of different kinds at various levels, which makes it difficult to quantify the empirical evidence to support this TU hypothesis.

6. Conclusions

Building on Xiao's (2010, 2011) initial investigations, the present study has explored a further range of lexical and grammatical properties of translational Chinese on the basis of two balanced comparable corpora of native and translational Chinese. The results show that translational Chinese differs from native Chinese in terms of various lexical and grammatical properties including, for example, the use of high-frequency words and low-frequency words, mean word length, keywords, distribution of word classes such as pronouns and prepositions, idioms and major punctuation marks at the lexical level, as well as mean sentence segment length, *bei* and *wei...suo* passives, disposal *ba* constructions, classifiers, aspect markers, structural auxiliaries, and modal particles at the grammatical level.

A reevaluation of the English-based TU hypotheses in the face of evidence from translational Chinese suggests that some (e.g. explicitation) are supported in Chinese while others are not fully supported (e.g. simplification) or even totally untenable (normalisation). More specifically, translational language is more explicit semantically, lexically, grammatically and logically. But simplification is not a pure, simple phenomenon in that translated texts may be simpler in some aspects but more complicated in others vis-à-vis comparable native texts. Translational language is a mixture of target language unique item under-representation and source language interference. As a result, translations from the same or similar source languages are likely to show more common properties. Given such commonalities, translated texts may appear more similar to each other than to non-translated native texts in the target language, but the extent of levelling out or convergence is dependent on the language pair involved in translation.

The present study has taken a comparable corpus approach to the investigation of the properties of translational language. While contrastive analyses based on monolingual comparable corpora can effectively uncover T-universals, and to some degree also S-universals, in translated texts, further research is required, on the basis of a parallel corpus, to establish the extent to which S-universals in translational language, such as explicitation and source language interference, are induced by the source language.

Acknowledgements

This research is partially supported by the Program for New Century Excellent Talents in University (grant ref. NCET-11-0460) by the Ministry of Education, China and by China's National Social Sciences Foundation Key Project "The Construction and Processing of Large Scale Chinese-English Parallel Corpus" (grant ref. 10ZD&127).

Bionotes

Richard Xiao is Lecturer in the Department of Linguistic and English Language at Lancaster University in the UK. His main research interests cover corpus linguistics, contrastive and translation studies of English and Chinese, and tense and aspect theory. In addition to dozens of journal articles, he

has published numerous books including *Aspect in Mandarin Chinese* (John Benjamins, 2004), *Corpus-Based Language Studies* (Routledge, 2006), *A Frequency Dictionary of Mandarin Chinese* (Routledge, 2009), *Using Corpora in Contrastive and Translation Studies* (Cambridge Scholars, 2010), *Corpus-Based Contrastive Studies of English and Chinese* (Routledge, 2010), and *Corpus-Based Studies of Translational Chinese in English-Chinese Translation* (Shanghai Jiao Tong University Press, 2012). Richard is a member of editorial boards for international journals including *Chinese Language and Discourse*, *Corpora*, *Foreign Language Learning Theory and Practice*, *Glossa*, *International Journal of Corpus Linguistics*, *Languages in Contrast*, and the *Corpus-Based Translation Studies* book series of Shanghai Jiao Tong University Press. Email: r.xiao@lancaster.ac.uk

Guangrong Dai is Associate Professor at Fujian University of Technology and a PhD candidate at the Department of English, University of Macau, China. His research interests include translation studies, corpus linguistics, contrastive language studies and designing software for automatic sentence alignment of Chinese/English parallel corpora. He has published over 30 journal articles and book chapters on corpus-based translation studies and contrastive language studies. Email: carldy75@gmail.com

Notes

1. Mona Baker used the term “universals” for the time but later retracted it and opted for a plain label “translational patterns and regularities” (cf. Kruger 2002: 99).
2. Other TU hypotheses that have been proposed in the literature but are not investigated in this study include examples such as sanitisation, i.e. translated texts are supposed to be “somewhat ‘sanitised’ versions of the original” (Kenny 1998: 515), and untypical collocation (Mauranen 2000).
3. Interjections, onomatopoeias, character strings and punctuation marks are excluded from analysis.
4. This example is taken from a bilingual magazine (www.taiwan-panorama.com).
5. In this study, the number of sentence segments is equivalent to the sum of the sentence number and the number of commas because sentence segments in Chinese are separated by commas.
6. It can also be used as a modal particle in the emphatic structure 是...的 *shi...de*.
7. The word can also be used as an archaic form of the third person pronoun in modern Chinese.
8. We are obliged to Naixing Wei (personal communication on 13 January 2013) for pointing out an alternative theory. The Chinese structural auxiliary 的 *de* started to be used much more frequently than before during and after the New Culture Movement in 1919, which marks the transition from classic Chinese to vernacular Chinese, because Hu Shi (leader of the movement) and colleagues regarded it as the Chinese equivalent of the English preposition *of*. In their writings and translations, 的 *de* was a very prominent structural element, which heavily influenced contemporary Chinese sentence structure in this particular aspect.

References

- 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* (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: Studies in Language Engineering in Honour of Juan C. Sager* (pp. 175–187). Amsterdam: John Benjamins.
- Baker, Mona. 2004. A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics*, 9(2). 167–193.

- Baker, Mona. 2007. Patterns of idiomaticity in translated vs. non-translated text. *Belgian Journal of Linguistics*, 21. 11–21.
- Baumgarten, Nicole, Meyer, Bernd and Özçetin, Demet. 2008. Explicitness in translation and interpreting: A critical review and some empirical evidence (of an elusive concept). *Across Languages and Cultures*, 9(2). 177–203.
- Becher, Viktor. 2010. Abandoning the notion of ‘translation-inherent’ explicitation: Against a dogma of translation studies. *Across Languages and Cultures*, 11(1). 1–28.
- Blum-Kulka, Shoshana. 1986. Shifts of cohesion and coherence in translation. In Juliane House and Shoshana Blum-Kulka (eds.), *Interlingual and Intercultural Communication* (pp. 17–35). Tübingen: Narr.
- Blum-Kulka, Shoshana and Levenston, Eddie. 1983. Universals of lexical simplification. In Claus Faerch and Gabriele Kasper (eds.), *Strategies in Interlanguage Communication* (pp. 119–139). London: Longman.
- Chen, Hsin-Hsi. 1994. The contextual analysis of Chinese sentences with punctuation marks. *Literary and Linguistic Computing*, 9(4). 281–289.
- Cheong, Ho-Jeong. 2006. Target text contraction in English-into-Korean translations: A contradiction of presumed translation universals? *Meta*, 51(2). 343–367.
- Chesterman, Andrew. 2004. Beyond the particular. In Anna Mauranen and Pekka Kujamäki (eds.), *Translation Universals: Do They Exist?* (pp. 33–49). Amsterdam: John Benjamins.
- Chesterman, Anna. 2007. What is a unique item? In Yves Gambier, Miriam Shlesinger and Radegundis Stolze (eds.), *Doubts and Directions in Translation Studies* (pp. 3–13). Amsterdam: John Benjamins.
- Dai, Guangrong and Xiao, Richard. 2011. ‘Source language shining through’ in translational language: A corpus-based study of Chinese translation of English passives. *Translation Quarterly*, 62. 85–107.
- Duff, Alan. 1981. *The Third Language: Recurrent Problems of Translation into English*. Oxford: Pergamon.
- Frawley, William. 1984. Prolegomenon to a theory of translation. In William Frawley (ed.), *Translation: Literary, Linguistic and Philosophical Perspectives* (pp. 159–175). London: Associated University Press.
- Hansen, Silvia and Teich, Elke. 2001. Multi-layer analysis of translation corpora: methodological issues and practical implications. In N. I. D. Cristea, D. Marcu and M. Poesio (eds.), *Proceedings of EUROLAN 2001 Workshop on Multi-layer Corpus-based Analysis* (pp. 44–55). Iasi.
- House, Juliane. 2004. Explicitness in discourse across languages. In Juliane House, Werner Koller and Klaus Schubert (eds.), *Neue Perspektiven in der Übersetzungs- und Dolmetschwissenschaft* (pp. 185–208). Bochum: AKS.
- House, Juliane. 2008. Beyond intervention: Universals in translation. *Trans-kom*, 1(1). 6–19.
- Hundt, Marianne, Sand, Andrea and Siemund, Rainer. 1998. *Manual of Information to Accompany the Freiburg-LOB Corpus of British English*. Freiburg: University of Freiburg.
- Jantunen, Jarmo Harri. 2004. Untypical patterns in translations: Issues on corpus methodology and synonymy. In Margaret Rogers and Gunilla Anderman (eds.), *Incorporating Corpora. The Linguist and the Translator* (pp. 101–126). Clevedon: Multilingual Matters.
- Kenny, Dorothy. 1998. Creatures of habit? What translators usually do with words. *Meta*, 43(4). 515–523.
- Kenny, Dorothy. 2001. *Lexis and Creativity in Translation. A Corpus-Based Study*. Manchester: St. Jerome Publishing.
- Klaudy, Kinga. 2009. Explicitation. In Mona Baker and Gabriela Saldanha (eds.), *Routledge Encyclopaedia of Translation Studies* (pp. 80–85). London: Routledge.

- Kruger, Alet. 2002. Corpus-based translation research: Its development and implications for general, literary and Bible translation. *Acta Theologica Supplementum*, 2. 70–106.
- Laviosa, Sara. 1998a. The corpus-based approach: A new paradigm in translation studies. *Meta*, 43(4). 474–479.
- Laviosa, Sara. 1998b. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, 43(4). 557–570.
- Laviosa, Sara. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi.
- Laviosa, Sara. 2009. Universals. In Mona Baker and Gabriela Saldanha (eds.), *Routledge Encyclopaedia of Translation Studies* (pp. 306–310). London: Routledge.
- Laviosa-Braithwaite, Sara. 1996. *The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation*. PhD Thesis. University of Manchester.
- Liu, Miqin. 1991. *Han Ying Duibi Yanjiu yu Fanyi (Contrastive of Chinese and English and Translation)*. Nanchang: Jiangxi Educational Press.
- Lü, Shuxiang. and Zhu, Dexi. 1979. *Yufa Xiuci Jianghua (Talks on Grammar and Rhetoric)*. Beijing: Chinese Youth Press.
- Malmkjær, Kirsten. 1997. Punctuation in Hans Christian Andersen's stories and in their translations into English. In Fernando Poyatos (ed.), *Nonverbal Communication and Translation: New Perspectives and Challenges in Literature, Interpretation and the Media* (pp. 151–162). Amsterdam: John Benjamins.
- Malmkjær, Kirsten. 2005. Norms and nature in Translation Studies. *SYNAPS*, 16. 13–19
- Mauranen, Anna. 1998. Käännösuomi ja kääntämisen universaalit: tutkimus korpusai-neistolla (Translated Finnish and Translation Universals: A Corpus Study). MS. Savonlinna School of Translation Studies.
- Mauranen, Anna. 2000. Strange strings in translated language: A study on corpora. In Maeve Olohan (ed.), *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects* (pp. 119–141). Manchester: St. Jerome Publishing.
- Mauranen, Anna. 2008. Universal tendencies in translation. In Gunilla Anderman and Margaret Rogers (eds.), *Incorporating Corpora. The Linguist and the Translator* (pp. 32–48). Clevedon: Multilingual Matters.
- Mauranen, Anna. and Kujamäki, Pekka. 2004. *Translation Universals: Do They Exist?* Amsterdam: John Benjamins.
- May, Rachel. 1997. Sensible elocution. How translation works in & upon punctuation. *The Translator*, 3(1). 1–20.
- 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.
- McEnery, Tony. and Xiao, Richard. 2008. Parallel and comparable corpora: What is happening? In Gunilla Anderman and Margaret Rogers (eds.), *Incorporating Corpora. The Linguist and the Translator* (pp. 18–31). Clevedon: Multilingual Matters.
- Nida, Eugene. 1964. *Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: Brill.
- Olohan, Maeve. 2001. Spelling out the optionals in translation: A corpus study. *UCREL Technical Papers 13* (pp. 423–432). Lancaster: UCREL, Lancaster University.
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. London: Routledge.
- Olohan, Maeve. and Baker, Mona. 2000. Reporting *that* in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures*, 1(2). 141–158.

- Pym, Anthony. 2008. On Toury's laws of how translators translate. In Anthony Pym, Miriam Shlesinger and Daniel Simeoni (eds.), *Beyond Descriptive Translation Studies: Investigations in Homage to Gideon Toury* (pp. 311–328). Amsterdam: John Benjamins.
- Reiss, Katharina. 1971. *Möglichkeiten und Grenzen der Übersetzungskritik: Kategorien und Kriterien für eine sachgerechte Beurteilung von Übersetzungen*. München: Max Heuber Verlag.
- Shlesinger, Miriam. 1989. Simultaneous interpretation as a factor in effecting shifts in the position of texts on the oral-literary continuum. MA thesis. Tel Aviv University.
- Sinclair, John McH. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, Michael. 1996. *Text and Corpus Analysis. Computer-assisted Studies of Language and Culture*. London: Blackwell.
- Teich, Elke. 2003. *Cross-linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Berlin: Mouton de Gruyter.
- Tirkkonen-Condit, Sonja. 2002. 'Translationese' – a myth or an empirical fact? A study into the linguistic identifiability of translated language. *Target*, 14(2). 207–220.
- Tirkkonen-Condit, Sonja. 2004. Unique items – over- or under-represented in translated language? In Anna Mauranen and Pekka Kujamäki (eds.), *Translation Universals: Do They Exist?* (pp. 177–184). Amsterdam: John Benjamins.
- Toury, Gideon. 1979. Interlanguage and its manifestations in translation. *Meta*, 24(2). 223–231.
- Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.
- Toury, G. 2004. Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals? In Anna Mauranen and Pekka Kujamäki (eds.), *Translation Universals: Do They Exist?* (pp. 15–32). Amsterdam: John Benjamins.
- Tymoczko, Maria. 1998. Computerised corpora and the future of Translation Studies. *Meta*, 43(4). 652–660.
- Vanderauwera, Roa. 1985. *Dutch Novels Translated into English: The Transformation of a "Minority" Literature*. Amsterdam: Rodopi.
- Vinay, Jean Paul. and Darbelnet, Jean Louis. 1958. *Stylistique Comparée du Français et de l'Anglais*. Paris: Didier-Harrap.
- Wang, Kefei. 2005. Fanyi zhong de yin yu xian (Implication and explicitation in translation). *Foreign Language Teaching and Research*, 2005(4). 303–307.
- Wang, Kefei. and Qin, Hongwu. 2010. A parallel corpus-based study of translational Chinese. In Richard Zhonghua Xiao (ed.), *Using Corpora in Contrastive and Translation Studies* (pp. 164–181). Newcastle: Cambridge Scholars Publishing.
- Wang, Li. 1985. *Zhongguo Xiandai Yufa (Modern Chinese Grammar)*. Beijing. Commercial Press.
- Xiao, Richard. 2010. How different is translated Chinese from native Chinese. *International Journal of Corpus Linguistics*, 15(1). 5–35.
- Xiao, Richard. 2011. Word clusters and reformulation markers in Chinese and English: Implications for translation universal hypotheses. *Languages in Contrast*, 11(1). 145–171.
- Xiao, Richard. 2012. *Ying Han Fanyi zhong de Hanyu Yiwen Yuliaoku Yanjiu (Corpus-based Studies of Translational Chinese in English-Chinese Translation)*. Shanghai: Shanghai Jiao Tong University Press.
- Xiao, Richard and Dai, Guangrong. 2010. Hanyu yiwen zhong xiyu yu cicu de shiyong tezheng: Jiyu yuliaoku de yanjiu (A corpus-based study of idioms and word clusters in translated Chinese). *Foreign Language Research*, 2010(3). 79–86.
- Xiao, Richard, He, Lianzhen and Yue, Ming. 2010. In pursuit of the third code: Using the ZJU Corpus of Translational Chinese in Translation Studies. In Richard Xiao (ed.), *Using Corpora in Contrastive and Translation Studies* (pp. 182–214). Newcastle: Cambridge Scholars Publishing.

- Xiao, Richard and McEnery, Tony. 2005. Two approaches to genre analysis: Three genres in modern American English. *Journal of English Linguistics*, 33(1). 62–82.
- Xiao Richard, McEnery, Tony and Qian, Yufang. 2006. Passive constructions in English and Chinese. *Languages in Contrast*, 6(1). 109–149.
- Xiao, Richard and Yue, Ming. 2009. Using corpora in translation studies: The state of the art. In Paul Baker (ed.), *Contemporary Corpus Linguistics* (pp. 237–262). London: Continuum.