Contents

KANGLONG LIU

6 Investigating corpus-assisted translation teaching: A pilot study

This contribution investigates how and the extent to which parallel corpora affect students' translation quality. The study presented is based on a large-scale web-based Chinese–English parallel corpus, that is, the Hong Kong Parallel cum Comparable Corpus (HKPCC), which comprises three text types (legal documents, news texts, parliament proceedings). The corpus has more than one hundred million Chinese characters with corresponding parallel English texts aligned at the sentence level. Since the study was set up as an empirical comparative study, a control group and an experimental group were used to test how students would perform translating in a corpus-assisted environment. Both groups were given the same piece of Chinese news text, which they were required to translate into English. The control group used conventional resources (i.e. monolingual and bilingual dictionaries, thesauri) for the translation assignment and the experimental group used a parallel corpus (i.e. an HKPCC excerpt). Data analysis of the student translations shows that the experimental group performed better in a number of aspects, including collocations, phraseology, spelling, terminology and word choice. Holistic scores given by two independent examiners also show that the experimental group produced better translation output than the control group. Based on the findings of the study, it is argued that parallel corpora are a useful resource for both translation teachers and translation students. In addition, it is argued that the corpus-assisted approach can be an innovative and effective means to complement the traditional translation teaching/learning approaches

Keywords: Empirical studies, parallel corpus, translation experiment, translation teaching

Introduction

As suggested by its name, corpus-assisted translation pedagogy is placed within a triangle formed by at least three distinct but not discrete disciplines: (1) corpus linguistics, (2) translation theory and (3) educational theory. The interdisciplinary nature of a corpus-assisted approach to translation teaching more or less dictates that a thorough investigation should take into consideration the influences and theories from these three areas. In the past decade, an exponential increase in literature output advocating the adoption of corpora for translation teaching has been witnessed. However, in contrast to a vast number of self-justifying papers, longitudinal and experimental research in this area is relatively rare. Empirical research into corpus-assisted translation teaching is important because a better understanding of the nature and process corpus-assisted translation will greatly inform pedagogy.

Corpus and translation pedagogy: An overview

In the field of translation teaching, the potential of corpora for informing translation teaching is also recognised by researchers. As Hunston (2002: 123) observes, research into corpora and translation tends to focus on two areas: (1) practical features and (2) theoretical features. In fact, the unidirectional influence of descriptive and theoretical branches on the practical branch of translation studies is the general trend of translation studies. Traditionally, second language learning and translation are treated as two inseparable activities since translation involves the mediation of two different languages. This is particularly the case when translation takes place from a person's mother tongue into a foreign language. For this reason, Bernardini (2004: 97) argues the following:

> Corpora have an important role to play in the education of translators, first as translation aids, as testified in the literature, secondly as sources of learning activities

> and of knowledge about the language, and thirdly and more importantly perhaps, as instruments through which approaches to language teaching and to translation teaching can be integrated into a coherent whole, with common aims and methods specific to this pedagogic setting.

Bernardini's statement indicates that translation, mostly a language-based activity, is closely related with a person's good command of both a source language (SL) and a target language (TL). In this respect, corpora are believed to be useful because they help translators to better understand languages and to improve the command of languages. Or, as Aston (1999: 292) puts it:

> [B]y drawing attention to the different ways expressions are typically used and with what frequencies, corpora can make learners more sensitive to issues of phraseology, register and frequency, which are poorly documented by other tools.

To a certain extent, the road to improving students' language competence (especially second-language competence) is where corpus-based language and translation pedagogy intersect. In this respect, even a monolingual corpus can prove helpful.

In the past decade, different types of corpora have been recommended and used for translation-teaching purposes. The most common types are monolingual corpora, which are further divided into SL monolingual corpus and TL monolingual corpus (Coffey, 2002). In translation-teaching practice, TL monolingual corpora are more popular among researchers since students translate into the target language. For instance, Bowker (1998) used a specialised monolingual corpus and discovered that corpus-aided translations were of a higher quality (in respect to subject field understanding, correct term choice and idiomatic expression) than those translations for which conventional resources such as dictionaries and encyclopedia had been used. Bowker (2000) further demonstrated the advantages of corpus use as a valuable resource for translators. By comparing two translations, one done using conventional resources and one done using a specialised corpus, Bowker (2000: 47) came to the conclusion that '[c]orpora are such valuable resources that translators would be remiss not to take advantage of what they have to offer'. Other researchers who have demonstrated the

usefulness of monolingual corpora for translation-teaching purposes include Bowker and Pearson (2002), Stewart (2000) and Wilkinson (2005). In 2008, a group of professional translators in Spain (Maher et al., 2008), argued – from a professional translators' perspective – that a corpus-guided approach, that is, using a TL monolingual specialised corpus, can provide translators with a faster and more economical way to master the terminology from specialised field than other means. Their views, arising from actual translation practice instead of academia, are worth noting in this respect.

Comparable corpora are also quite popular among teacher–researchers since they are also relatively easy to construct. Most comparable corpora contain specialised texts, that is, texts belonging to genres or domains that are sociolinguistically similar in each of the cultures involved (in terms of participation framework, function, and topic) (Aston, 1999: 291). Generally speaking, a comparable corpus, which indeed consists of one SL monolingual corpus and one TL monolingual corpus, has all the advantages and attributes of monolingual corpora. Kübler (2003: 41), combining different types of corpora and the Web for training terminologies, claimed that the use of 'comparable corpora in LSPs helps to overcome problems of "artificiality" in parallel corpora.'[1] By using a comparable corpus of English and Italian newspaper texts, Zanettin (2001) reported on a study in which a group of undergraduate students were asked to translate part of a newspaper article from Italian into English. He concluded that '[u]sing comparable corpora and concordancing software as aids in translation activities can help learners gain insights into the language and the cultures involved and develop their reading writing skills' (Zanettin, 2001: 193–194). Other studies dealing wholly or partially with comparable corpora include Aston (1999) and Maia (2003). However, most of these studies are seldom based on longitudinal systematic studies so their claims are not extremely forceful.

For educational purposes, parallel corpora are less studied in comparison to other types of corpora. This is partly due to a lack of accessible

1 The concept of a parallel corpus is a controversial one because the term is used by some researchers to refer to comparable corpora. It is used in this chapter to refer to a bilingual corpus that holds both original texts (source texts, STs) and their corresponding translated texts (target texts, TTs).

bilingual concordancing software and a scarcity of parallel texts. The term *parallel corpus* is used here to designate a collection of texts in language A and their correspondent translations into language B (Baker, 1995; Zanettin, 2001). A parallel corpus is often aligned at the sentence level to allow a concurrent display of the source text and its corresponding translations through keyword searches.

The exploitation of parallel corpora for pedagogical purposes was initially mostly connected with terminology extraction and teaching (Danielsson and Ridings, 2000; Pearson, 2000; Meyer et al., 2000). At the same time, the exploitation was also recommended for teaching languages because it can provide information on collocations, lexical polysemy and phrasal patterns (Barlow, 2000) and for teaching translation because it can extract bilingual information of collocations and idioms (Peters et al., 2000). Pearson (2003), using a small parallel corpus of popular science articles translated from English into French, demonstrated through the translation of university names that translation strategies differ in that some of these names were directly translated, some were not translated at all, and a few were translated by other means. Pearson (2003: 23) argued that by examining the translation strategies of previous translators, students can use the corpus evidence to 'draw up their own translation guidelines'.

However, it should be noted that most of the studies reported above were based on small-size parallel corpora and that the findings were often based on the researchers' own intuitive analysis and reasoning instead of on empirical studies or teaching experiments. This is mostly due to the scarcity of readily available parallel corpora.

Hong Kong Parallel cum Comparable Corpus

The Hong Kong Parallel cum Comparable Corpus (HKPCC) comprises four sub-corpora, which are labelled News, Hansards, Laws and CN News respectively. The News sub-corpus contains press releases issued by the

Government of the Hong Kong Special Administrative Region (HKSAR) in the period July 1997–October 2003. The Hansards sub-corpus contains excerpts from the Official Record of Proceedings (Hansards) of the Legislative Council of the HKSAR in the period October 1985–April 2003. The Laws sub-corpus contains only Hong Kong statute laws in English and Chinese, constitutional instruments, national laws and other relevant instruments published up to 2,000. The CN News sub-corpus is different since the data are coded in simplified Chinese instead of traditional Chinese and the sources are the Xinhua News Service and the AFP News service. This sub-corpus contains 1,001 news stories published in the period July 2002–August 2004.

Methodology

Views on the future

The current study was designed to investigate the following two research questions: (1) In comparison to conventional reference tools such as dictionaries, will the use of parallel corpora contribute to better translation quality when student translators are translating from their mother tongue (L1) into a foreign language (L2)? And (2) in comparison with the use of dictionaries, in what ways will students' translations change when they are given the opportunity to translate in a corpus-assisted manner?

Participants

The research presented here was carried out at the South China Normal University (SCNU). The university is one of the key universities in southern China. Forty-four students (23 in the control group and 21 in the experimental group), aged between 18 and 22 years of age, participated in the

study. The students who took part in the study were third-year undergraduates majoring in English (with a business translation focus).

Experimental design

Pretest

The pretest consisted of a translation task, in which students were required to translate a piece of news from Chinese into English. The text was taken from a Hong Kong newspaper dealt with the difficulties that mainland students experience when trying to find a job in Hong Kong. The text contained no specialised knowledge that might hamper the students' comprehension. Students were told to finish the translation within the hour. They were required to write down their translation (using a pen and a pencil) in the classroom. They were given access to any kind of monolingual or bilingual dictionaries during the translation task. Two dictionaries were provided to the students before the pretest took place: (1) the Oxford Learner's Dictionary (monolingual) and (2) the Modern Chinese–English Dictionary (bilingual).

Posttests

Both an inter-subject and intra-subject experimental design were used in the posttests. The inter-subject experimental design was aimed at comparing students' translation performance in the experiment group with translation performance in the control group. The intra-subject experiment design was aimed at comparing the translation performance of the experimental group in two situations, that is, translation with dictionaries (henceforth EG1) and translation with parallel corpora (henceforth EG2). In an attempt to control the variable *practice effect*, that is, improved translation performance in the experimental group as a result of repeated practice, the second translation task, in which students were asked to translate with parallel corpora, was arranged two weeks after the first translation task was completed. The purpose was to examine the extent to which corpus use might affect students' translation quality.

Data analysis

For the three translation tests (i.e. one pretest and two posttests), both quantitative and qualitative analyses were used to study the students' translated texts. The quantitative analysis was based on the scores given by two external examiners while the qualitative analysis consisted mainly of a qualitative examination of the students' translations.

The pretest and posttests were marked by two external examiners other than the current researcher to ensure objective evaluation of the students' work. The set of assessment criteria was adapted from the criteria proposed by Kiraly (1995: 83). However, a 10-point assessment scale, instead of Kiraly's original 5-point scale was adopted. In so doing, the examiners were given more room to judge the quality of different translations. The assessment scale is provided in Table 6.1.

Table 6.1: The 10-point assessment scale for the translated texts

Scale rank	Description
1–2	This is a totally unacceptable translation.
3–4	This is a poor translation. It would require major improvements before being submitted to an employer.
5–6	This translation is marginally adequate. It has several errors and would require a moderate amount of work to prepare it for submission to an employer.
7–8	This is essentially a good translation. It does have some minor errors, but they could be eliminated quite easily.
9–10	This is a very good translation. It contains no errors with respect to the norms of the target language and it is a functionally acceptable translation of the source text.

Results

Pretest

In order to investigate the inter-rater reliability between the two sets of scores given by the two external examiners, Pearson's correlation coefficient was computed. The result was significant, with a positive correlation between the two sets of scores, $r = .678$, $p = <.001$. An independent-samples *t*-test was used to determine whether there was a difference in mean pretest scores between the two groups. The result shows that the *t*-test was insignificant, $t = .407$, $df = 42$, $p = .686$. On average, the two groups appeared quite similar in their translation performance when using dictionaries as reference resources.

Posttests

Quantitative analysis of holistic scores

The inter-subject analysis was conducted by comparing the translation scores from the control group with the scores from the experimental group. The mean scores of these two sets of translations were calculated using an independent-samples *t*-test. The *t*-test was significant ($t = -7.086$, $df = 42$, $p = <.001$).

The statistical analysis of the scores shows that the experimental group produced better translations than the control group. The hypothesis that students using corpora can produce better translations than students using dictionary resources was affirmed in this inter-subject translation experiment.

A paired samples *t*-test was used to compare translations done with dictionary resources and translations done with the parallel corpus within the same group of participants. The result indicated that there was a significant difference in the scores between translation done with dictionary resources ($M = 3.786$, $SD = 1.3562$) and translations done with the corpus

($M = 5.548$, $SD = .8501$), $t = -5.116$, $p = <.001$. The results suggest that the parallel corpus had a positive effect on students' translations. Next, some features of the translations will be compared and analysed to provide a full picture of corpus-assisted translation conditions.

Spellings

On average, students in the control group made most spelling errors (6.74), followed by EG1 (5.05) and EG2 (0.43). This finding suggests that the use of word-processing software in conjunction with corpora proves to be an effective tool for doing translations in terms of spelling-error reduction. Students using corpora in conjunction with word-processing software made fewer spelling errors in the second translation than they did in their first translation, for which they used pen and paper. However, such an improvement cannot be completely attributed to the use of corpora. It might be more related to the use of word-processing software, which has a function that highlights spelling errors. It is hard to know how much a corpus, as opposed to a word-processor, influence this variable. A precise answer to this question would require considerable research into the translation process and corpus-assisted translation teaching.

Word choice

The analysis of word choice was conducted on two sets of translations (EG1 and EG2). There were two considerations for this decision. First, it is believed that an intra-subject analysis should yield more findings since the subjects are the same. Second, data analysis based on the comparison of these two sets of translations can be made concise and easy to follow. The software used for analysing this feature was Wmatrix, which is the Web interface to the USAS and CLAWS corpus-annotation tools developed by Paul Rayson. Through a comparison of keywords in the two datasets of students' translations, the overused words or word strings of each dataset were retrieved. Such a comparison can help to quickly identify the different features resulting from the two phases of translations.

In Table 6.2, the top 20 overused words (or word strings) by EG1 (as compared with EG2) are listed. The first column is the serial number of

the items. The second column shows the words or word string. The third column provides the number of occurrences of the words or word strings in the corpus, and the percentage is shown in the fourth column. Columns five and six are the same as the previous two, but this they provide the occurrences of the item in the other dataset (i.e. EG2). The last column is the log-likelihood value. The log-likelihood (LL) test can be used for corpus comparison. In the current study, log-likelihood is used to test the overused items between the two sets of translations.

Table 6.2: Top 20 overused items by EG1 (as compared with EG2)

No.	Item	EG1	%	EG2	%	LL
1	Report	17	0.31	0	0.00	23.73
2	administrative	25	0.45	2	0.04	23.40
3	Banks	33	0.60	8	0.14	16.61
4	Organisations	15	0.27	1	0.02	14.84
5	administration	16	0.29	2	0.04	12.53
6	Hong Kong	14	0.25	2	0.04	10.24
7	Finance	14	0.25	2	0.04	10.24
8	Of	232	4.19	171	3.06	9.87
9	goal	11	0.20	1	0.02	9.85
10	department	7	0.13	0	0.00	9.77
11	administrator	5	0.09	0	0.00	6.98
12	director	5	0.09	0	0.00	6.98
13	target	5	0.09	0	0.00	6.98
14	except	8	0.14	1	0.02	6.27
15	comprehensive	10	0.18	2	0.04	5.90
16	excluding	4	0.07	0	0.00	5.58
17	government	4	0.07	0	0.00	5.58
18	grave	4	0.07	0	0.00	5.58
19	Minister	4	0.07	0	0.00	5.58
20	server	4	0.07	0	0.00	5.58

Table 6.3 presents the top 20 overused words (or word strings) by EG2 (as compared with EG1). Through a comparison of Tables 6.2 and 6.3, we can see that students used different lexical terms in translating the same text. In addition, the disparity of lexical use in the two datasets indicates that parallel corpora have a positive influence on students' translations. Overall, lexical usage (in translated texts) resulting from corpus use is more sophisticated and consistent with the target norms than dictionary-based translations. Some major features will be discussed in the following sections.

Table 6.3: Top 20 overused items by EG2 (as compared with EG1)

No.	Item	EG1	%	EG2	%	LL
1	non-banking	18	0.32	0	0.00	24.78
2	policy	58	1.04	19	0.34	20.32
3	address	22	0.39	2	0.04	19.31
4	rescue	12	0.21	0	0.00	16.52
5	varied	10	0.18	0	0.00	13.77
6	's	9	0.16	0	0.00	12.39
7	institutions	39	0.70	14	0.25	12.03
8	throw	12	0.21	1	0.02	10.86
9	objective	13	0.23	2	0.04	8.91
10	Hong Kong	29	0.52	11	0.20	8.22
11	maintain	20	0.36	6	0.11	7.82
12	grim	15	0.27	4	0.07	6.68
13	confusion	13	0.23	3	0.05	6.64
14	restore	13	0.23	3	0.05	6.64
15	ourselves	12	0.21	3	0.05	5.70
16	ahead	4	0.07	0	0.00	5.51
17	amounts	4	0.07	0	0.00	5.51
18	peripheric	4	0.07	0	0.00	5.51
19	principal	4	0.07	0	0.00	5.51
20	worldwide	4	0.07	0	0.00	5.51

Terminology

When translating using dictionaries, students were not as accurate with respect to their choice in terminology. As can be seen in Table 6.2, the two words *report* and *administrative* top the list. The phrase *administrative report* was used by most students to translate 施政報告 (shi zheng bao gao) when they used dictionaries as reference resources. However, the correct term was *policy address*. The term was used by almost all the students using the HKPCC. When translating with the help of corpora, students were able to identify the correct term usage immediately by examining the co-occurrence of Chinese and English parallel texts in the corpus that they used. By contrast, when the students used the dictionaries as reference resources, some special terms were not found. The term *policy address* was a typical example.

The translation of the government posts 行政長官 (xing zheng zhang guan) and 財政司長 (cai zheng si zhang) was also a terminological challenge. When translating in a dictionary-based manner, a number of students used *administrator*, *director* and *minister* for the translations of these two terms. By contrast, most students used the correct terms – *Chief Executive* and *Financial Secretary* respectively – when they used the HKPCC. The terms provided in the dictionary do not distinguish between the practices in different countries in terms of addressing official titles. Students were clearly influenced by the official practice in China, where 部長 (bu zhang) or 司長 (si zhang) corresponds to *minister* or *director* instead of *secretary*. When translating with the help of the parallel corpus, students were able to retrieve the appropriate expressions after entering the Chinese terms into the corpus. Data analysis shows that 20 students in EG2 used the correct term (*Financial Secretary*) while only 5 in EG1 used the same expression. This indicates that corpus use – in comparison to dictionaries – can increase the accuracy rate for the translation of special terms.

Phraseology

From Table 6.3, it can be seen that the word string that tops the EG2 dataset is *non-banking*. Eighteen of the 21 students in the corpus group used the term *non-banking* as a translation for 銀行以外的 (yin hang yi wai de),

while in the dictionary group, such a fixed expression was not used at all by the students. Instead, students rendered the expression word-for-word, as is evidenced by the overuse of the words *banks* and *except* in the EG1 dataset. However, although paraphrasing is accepted as a proper technique in translation, the standard phrase *non-banking financial institutions* suits the context because of its conciseness and appropriateness. The use of standard phrases also adds to the coherence of the whole text. Below some examples taken from the students' translations that were investigated are provided.

EXAMPLE 1 (Subject 2–3)

a. Attentions must be paid to the financial institutions in occidental area, *with the exception to the banks*, to be alert to the new issues. (EG1)

b. Attention must be paid to the *non-banking financial institutions* in the occidental area to be alert to the new problems. (EG2)

EXAMPLE 2 (Subject 2–5)

a. Attention must be paid to whether there will be new problems facing other *financial organizations besides banks* in European and American areas. (EG1)

b. Attention must be paid to whether there will be new problems facing *non-banking financial institutions* in Europe and the United States. (EG2)

EXAMPLE 3 (Subject 2–17)

a. In Europe and America, the *financial insituations except banks* whether have new problems or not should be taken notice. (EG1)

b. In Europe and America, whether *the non-banking financial institutions* have new problems or not should be taken notice. (EG2)

The three examples above indicate that students in the dictionary group used *except*, *besides* and *excluding* to render the Chinese 以外的 (yi wai de). However, all students switched to the expression *non-banking*, a more standard expression to summarise the idea in the source text without any loss in meaning. The improvement of such a usage should clearly be attributed to the use of parallel corpus. To a certain extent, this finding also confirms

Bowker's study (1998), in which it was stated that corpus use can improve idiomatic expressions in students' translations since the two concepts of idiomatic expressions and standard phrases share certain similarities.

Collocation

Baker (1992: 14) discusses the notion of collocation and focuses on the restricted type, which she defines as 'semantically arbitrary restrictions which do not follow logically from the propositional meaning of a word'. In translation, collocation constitutes a major problem, especially when translators are not aware of the foreign-language or foreign-culture norms. In many cases, direct translation of certain elements often results in unnatural collocations in the target texts. In the current translation experiment, it was found that students using the corpus performed better in this area than students using dictionaries. Below, an analysis will conducted of the translations of three chunks of Chinese selected from the source text:

1. 外圍經濟環境 (wai wei jing ji huan jing),
2. 注入巨額資金 (zhu ru ju e zi jin) and
3. 恢復全球金融市場的秩序和信心 (hui fu quan qiu jin rong shi chang de zhi xu he xin xin).

In the translation experiment text, the term 外圍經濟環境 (wai wei jing ji huan jing) proved a difficult chunk for the students in their translation test. In the two datasets, most students rendered the term as either *peripheral economic environment* or *external economic environment*. The translation is focused on the use of the term 外圍 (wai wei), which can be directly translated as *peripheral*. However, when *peripheral* is used together with *economic environment*, the collocation *peripheral economic environment* does not seem to match the target-language norm. In order to retain the same meaning and make it compatible with the target-language norm, *external* is clearly a better choice. However, the translation provided for 外圍 by most dictionaries is *peripheral* and the dictionaries did not inform the students of its usage, and most students opted for *peripheral* instead of *external*. On the other hand, the students using the corpus were able to test their intuitions in the

corpus. It turned out that more students in the corpus group chose to use the term *external economic environment* when translating with the corpus.

In the translation test, the best translation for the term 注入巨額資金 (zhu ru ju e zi jin) was *inject a large sum of capital (into)*. However, most students using dictionaries chose to use other terms, such as *pour large amount of money* (Subject 2–1 in EG1), *invested heavily into* (Subject 2–4 in EG1), *infusing large amount of fund* (Subject 2–5 in EG1) and *put a great deal of capital a[c]tively in* (Subject 2–17 in EG1). These collocations were rather unnatural. As a matter of fact, many bilingual dictionaries give *inject* as the translation for 注入 (zhu ru), but the lack of context in dictionary resources seemed to deter the students from adopting the term. Instead, they tried to paraphrase the term and use other expressions. When translating using dictionaries, only four students used the term *inject* in the past-participle form *injected*. When they switched to using a corpus, eight students used the word *inject* or *injected*. Although the number increased by half, the overall percentage within the group was not terribly high (8 out of 21). Nevertheless, this example shows that corpus can be useful for improving the translation of certain collocations.

In the expression 恢復全球金融市場的秩序和信心 (hui fu quan qiu jin rong shi chang de zhi xu he xin xin), the translation is focused on the verb phrase 恢復 (hui fu), which can be directly translated as *recover*, *restore*, *resume*, etc. However, when the expression is used to collocate with *order* (秩序 zhi xu) and *confidence* (信心 xin xin), *restore* is clearly the best choice. In the EG1 dataset, students used a variety of other verbs such as *recover*, *regain* and *resume*. Only three students used the word *restore* when translating using dictionaries. The rest of the collocations (e.g. *maintain the order*, *recover the order*, *regain the rules*) do not seem to comply with the target-language norm. Other renderings include *recover* (eight instances) and *recovery* (one instance). This choice is clearly influenced by the context-detached translations provided by some bilingual dictionaries. On the other hand, students using the corpus clearly performed better in terms of rendering this term with a natural collocation. Sixteen students in the corpus group used the word *restore* or its inflected forms. In most cases, the word was used to collocate with *order* and *confidence* to form natural-sounding collocations.

This indicates that the translation of collocations in the corpus group was clearly better than the translation of collocations in the dictionary group.

Results

Two research questions were formulated and were used to set up the study being reported on in this chapter. These two questions were the following: (1) In comparison to conventional reference tools such as dictionaries, will the use of parallel corpora contribute to better translation quality when student translators are translating from their mother tongue (L1) into a foreign language (L2)? And (2) in comparison with the use of dictionaries, in what ways will students' translations change when they are given the opportunity to translate in a corpus-assisted manner?

In reference to the first research question, it can now be stated that the use of parallel corpora – in comparison with the use of conventional reference tools such as dictionaries – can indeed contribute to better translation quality when student translators are translating from their mother tongue (L1) into a foreign language (L2). The effectiveness of the parallel corpus for L1–L2 translation is proved by both quantitative and qualitative evidence obtained from students' translated texts. Nevertheless, the finding is limited to the research context which consists of the current study and includes variables such as the level of students, the type of corpus and the direction of translation (from Chinese into English). Because the parallel corpus was aligned at the sentence level, the source text corresponded with the target text on a sentence-to-sentence basis. In other words, semantic equivalence between source text and target text at the sentence level is a prerequisite for the establishment and application of parallel corpora, as is the case for the study presented here, which was based on the HKPCC. Non-correspondent translations would not be included in the parallel corpus. Consequently, parallel corpora take on a prescriptive element, which is non-existent in monolingual corpora and

comparable corpora. Interestingly enough, the prescriptive nature of parallel corpora is in line with translation teaching since the latter also imposes a certain degree of prescriptiveness on teachers and students. The use of the one-to-one sentence alignment design in the parallel corpus results in students' being able to easily identify solutions to translation problems once translation instances have been found in the corpus. Because of this, parallel corpora are useful for translation purposes, particularly in translation-training contexts in which students are not proficient in mediating between two languages and cultures. In most cases, students can find a number of instances of certain translation problems if corpora are big enough. By analysing these instances – which are recognised as translation solutions – with other translators, students appear able to judge for themselves and make their own choices. For L1–L2 translation, the use of parallel corpora can help students to overcome the limitations of language and cultural barriers since the translation solutions are provided in given contexts. These translation solutions by professional translators are conditioned by factors such as text types, translation purpose, readership, and they can be placed on continuum from formal to semantic correspondence. By examining and comparing the 'context likeliness' of corpus occurrences and their current translation problems, students are able to render better and more adequate translations than when they work without corpora. This is especially true for L1–L2 translation, where students are not generally equipped with the necessary cultural knowledge of an L2 text. The use of parallel corpora can fill the gaps in students' knowledge of foreign cultures, thus improving their translation quality when they are given such a tool.

The second research question was addressed by studying the translated texts that were produced in the two translation experiments. The parameters chosen to study the translated texts included the number of misspellings, word choice, terminology, phraseology and collocations. In general, it was found that students made improvement in these areas when translating in a corpus-assisted manner.

However, it is suspected that the effectiveness of parallel corpora is related to the size of translation unit. In other words, the effectiveness of parallel corpora decreases as the translation unit become bigger and vice

versa. The improvement was most obvious at the word level, which is evidenced by the more sophisticated vocabulary in the translation tests. The correct use of terminology and phraseology in the English translations also reinforces this point.

Concluding remarks

The research presented here has demonstrated the pedagogical value of using parallel corpora in actual translation teaching settings. Nonetheless, the research scope and design can be improved in a number of ways. For instance, some major findings from the current study are based on the analysis of corpus-assisted translation products rather than on the analysis of corpus-assisted translation processes, which could have added more insights into the research findings. Because translation products take the form of end-products that reflect few process-related cognitive moves in translation, such as changes, number of concordancing in the corpus, the degree of reliance on exact matches, the findings are limited. Future studies can use think-aloud protocol (TAP) methods or screen recording software to detect and to track the translation process, with a view to uncovering the unique cognitive patterns in corpus-assisted translation contexts. It is believed this type of research can be used to corroborate some of the findings in the study presented here.

Another area that might be interesting for future studies is the introduction of a different variable – in the form of literary texts – to the translation experiments. In corpus-assisted translation-teaching research, an overwhelming number of studies (Bowker, 1998, 2000; Corpas and Seghiri, 2009; Kübler, 2003; Maia, 2003), as it is with the current study, are focused on only non-literary texts. Consequently, the findings may not be applicable to all text types. This has also hampered any further discussions of, for example, the opposition creativity–conservatism in translation because non-literary texts allow little room for creativity on the part of translators. Therefore, translation experiments could be conducted with literary texts

by using a large-scale literary corpus to examine how students perform with such texts.

Lastly, on a technical note, future studies could be devoted to the development of more sophisticated corpus tools for translation teachers and researchers. As is suggested by some of the student participants, there is still much room for corpus improvement. Interdisciplinary collaboration between computer technicians and translation teachers could take place to make the corpus teaching platform more effective by including additional functions. After all, the success of corpus-assisted translation teaching should address both technological and academic issues, the former being the precinct of computer experts and the latter the area of translation teachers and researchers. Collaborative efforts could be made to investigate how to bring the best of both world (computer experts and translation researchers) together to benefit corpus-assisted translation teaching.

Bibliography

Aston, G. (1999). Corpus use and learning to translate. *Textus, 12*, 289–314.

Baker, M. (1992). *In Other Words: A Coursebook on Translation*. London: Routledge.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target, 7*(2), 223–243.

Barlow, M. (2000). Parallel texts in language teaching. In S. P. Botley, A. M. McEnery, and A. Wilson (eds), *Multilingual Corpora in Teaching and Research* (pp. 106–115). Amsterdam: Rodopi.

Bernardini, S. (2004). Corpus-aided language pedagogy for translator education. In K. Malmkjær (ed.), *Translation in Undergraduate Degree Programmes* (pp. 97–111). Amsterdam: John Benjamins.

Bowker, L. (1998). Using specialized monolingual native-language corpora as a translation resource: A pilot study. *Meta, 43*(4), 631–665.

Bowker, L. (2000). Towards a methodology for exploiting specialized target language corpora as translation resources. *International Journal of Corpus Linguistics, 5*(1), 17–52.

Bowker, L., and Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London: Routledge.

Coffey, S. (2002). Using a source language corpus in translator training. *inTRAlinea,* 5. <http://www.intralinea.org/specials/article/Using_a_Source_Language_Corpus_in_Translator_Training>.

Corpas, G., and Seghiri, M. (2009). Virtual corpora as documentation resources: Translating travel insurance documents (English–Spanish). In A. Beeby Lonsdale, P. Rodríguez Inés, and P. Sánchez-Gijón (eds), *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate* (pp. 75–107). Amsterdam: John Benjamins.

Danielsson, P., and Ridings, D. (2000). Corpus and terminology: Software for the translation program at Göteborgs Universitet. In S. P. Botley, A. M. McEnery, and A. Wilson (eds), *Multilingual Corpora in Teaching and Research* (pp. 65–72). Amsterdam: Rodopi.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Kiraly, D. C. (1995). *Pathways to Translation: Pedagogy and Process*. Kent, OH: Kent State University Press.

Kübler, N. (2003). Corpora and LSP translation. In F. Zanettin, S. Bernardini, and D. Stewart (eds), *Corpora in Translator Education* (pp. 25–42). Manchester: St. Jerome.

Maher, A., Waller, S., and Kerans, M. E. (2008). Acquiring or enhancing a translation specialism: The monolingual corpus-guided approach. *The Journal of Specialised Translation, 10*, 56–75.

Maia, B. (2003). Some languages are more equal than others: Training translators in terminology and information retrieval using comparable and parallel corpora. In F. Zanettin, S. Bernardini, and D. Stewart (eds), *Corpora in Translator Education* (pp. 55–70). Manchester: St. Jerome.

Meyer, R., Okurowski, M. E., and Hand, T. (2000). Using authentic corpora and language tools for adult-centred learning. In S. P. Botley, A. M. McEnery, and A. Wilson (eds), *Multilingual Corpora in Teaching and Research* (pp. 86–91). Amsterdam: Rodopi.

Pearson, J. (2000). Teaching terminology using electronic resource. In S. P. Botley, A. M. McEnery, and A. Wilson (eds), *Multilingual Corpora in Teaching and Research* (pp. 92–105). Amsterdam: Rodopi.

Pearson, J. (2003). Using parallel texts in the translator training environment. In F. Zanettin, S. Bernardini, and D. Stewart (eds), *Corpora in Translator Education* (pp. 15–24). Manchester: St. Jerome.

Peters, C., Picchi, E., and Biagini, L. (2000). Parallel and comparable bilingual corpora in language teaching and learning. In S. P. Botley, A. M. McEnery, and A. Wilson (eds), *Multilingual Corpora in Teaching and Research* (pp. 73–85). Amsterdam: Rodopi.

Stewart, D. (2000). Conventionality, creativity, and translated text: The implications of electronic corpora in translation'. In M. Olohan (ed.), *Intercultural Faultlines: Research Models in Translation Studies* (pp. 73–91). Manchester: St. Jerome.

Wilkinson, M. (2005). Using a specialized corpus to improve translation quality. *Translation Journal, 9*(3). <http://accurapid.com/journal/33corpus.htm>.

Zanettin, F. (2001). Swimming in words: Corpora, translation and language learning. In G. Aston (ed.), *Learning with Corpora* (pp. 177–197). Bologna: CLUEB.