Specialized Parallel Corpus-based Military Translator Training

Xiangbing Wang

National University of Defense Technology, China

Abstract: The influence of corpus on translator competence has become a hot issue in the corpus-based translational studies. This paper, starting from the empirical study of a mini DIY parallel corpus (400,000 words) of military texts, makes a primary probe into the influence of specialized parallel corpus on translator competence and reaches the conclusion that specialized parallel corpus can enhance translator's translation speed subject-field understanding significantly and promote translators' competencies on foreign language as well as native language in a tangible way. A special experiment of corpus-based military translation training was carried out to explore the relationship between a specialized parallel corpus and translator competence. The parameters set for measuring translator competence in the experiment are: translation speed, foreign language competence, native language competence and professional knowledge.

1. Introduction

The combination of corpus and translation studies happened at the end of 20th century. After decades of development, corpus-based translation studies is enlarged to have extended to all kinds of fields. With the development of computer technology, corpora

with different functions and scales are already built. According to the function and nature, there are general corpora, specialized corpora, monitor corpora, oral corpora, student English corpora, parallel corpora, etc.

Parallel corpora imply to input the completely equivalent texts in two languages into a computer and find out the equivalent relationships between them through comparison and analysis. (Yang 2002:29) A parallel corpus is most helpful to translation studies because it can align the source language and the target language on textual level, sentence level or word level and display both before translators.

Whereas specialized corpora are built for special purposes, for example, a specialized corpus for researching old English. These corpora for particular purposes can play important roles in analyzing and dealing with the linguistic and stylistic features of the texts in or provide background knowledge of a particular field. So far, a lot of specialized corpora of diversified fields are already built and in use, such as legal corpus, medical corpus, among many others.

The two above-mentioned corpora have their respective advantages and the specialized parallel corpora[①] which have combined the two kinds of corpora are especially helpful to translation studies. This paper is going to explore preliminarily how specialized parallel corpora can play their roles in military translator training.

2. The Corpus Used

Military corpus-based study is comparatively underdeveloped and cut off because of

[①] This term is temporarily used in this paper since there is no unanimous acceptance of this term yet in international academia.

the specialty of the field concerned, but its importance is undeniable and there are also diversified and multifunctional military corpora (Liang et al 2008). PLA University of Foreign Languages has built a military corpus. The author of this paper has recently finished an academic project based on a military corpus which is specially designed and built by the project team. The language material used in this corpus comes from authoritative military departments and the translation is publicly published. This corpus can be rated as a small-scale self-built military translation parallel corpus.

In the process of building the corpus, the biggest challenge is not technology, nor manpower, but data collecting. It's very difficult to find non-confidential bilingually corresponding texts which reflect perfectly the features of military translation; or at least the translated texts are produced by professional military translators to reflect the authoritativeness of the corpus. All these minute details make the text collecting work for a parallel corpus more difficult. As a result, the scale of the present corpus is somewhat confined[②], but it's big enough to carry out some small-scale experiments.

The alignment of this corpus is based on sentence level. The program principles are simple: first, run a self-made program to separate the parallel texts into one-sentence paragraphs and number each paragraph; second, do manual proofreading; finally, run the program again to delete the numbers and process the parallel texts further, for example, to add <S> at the beginning of every paragraph and </S> at the end of every paragraph and add text titles as <CH_TITLE>…</CH_TITLE> or <EN-TITLE>… </EN_TITLE>, etc. Besides, the Chinese parsing software used by the corpus is

[②] The corpus is open. It is still being built and the scale can become larger and larger to cater for various corpus-based military purposes.

ICTCLAS which is developed by Chinese Academy of Sciences; the corpus analysis software is Paraconc which can be downloaded from the following website: http://www.athel.com/para.html.

Obviously, this corpus represents the features of both parallel corpora and specialized corpora. As a military translation parallel corpus, its biggest advantages are to provide authoritative bilingual translation in military fields, related military background knowledge and reference of stylistic features of military English and military Chinese, etc.

3. Methodology[③]

The experiment in this research has set two parameters: translation speed and translation quality, which are two general parameters to measure translator competency. However, translation speed is easy to be measured, but translation quality is difficult to be quantified. Liu Qun (2000) in his article *On the Quality Standard and Knowledge Structure of Scientific Translation* proposes that scientific translation's knowledge structure includes three aspects: foreign language, native language and subject-field knowledge. Sykes (1989: 35-39) points out that a good translator must possess three competencies: thorough understanding of the subject-field knowledge in the source text, a good commanding of target language (usually the translator's mother tongue) and a good commanding to the source language (usually a foreign language). Accordingly, it is advisable to specify

[③] The methodology of the experiment in this research has referred to three articles: Bowker, 1998; May & Xu, 2002; Wei, 2006.

translation quality as foreign language, native language and subject-field knowledge in order to analyze the experiment data. Therefore, we have got the four parameters in this experiment to measure translator competency, i.e. translation speed, foreign language competency, native language competency and subject-field knowledge. So the purpose of this experiment can be specified as to investigate how the military translation parallel corpus influences the cadets' translation speed, foreign language competency, native language competency and subject-field knowledge. Besides, it's necessary to know how the students think of the corpus.

The subjects of this experiment are 24 volunteering undergraduate English majors who are going to be seniors soon in a Chinese military university. They have no experience with a corpus before the experiment, but they have already had experiences of doing military translation. English is their first foreign language and Chinese is their mother tongue. They have accumulated a certain amount of military background knowledge, but far from being experts.

The two source texts chosen for the experiment are respectively 299 words (text 1) and 307 words (text 2) long. Both texts are chosen from some specialized military field which may be quite difficult for these subjects. When the experiment time is set and the result is analyzed the above factors should be taken into consideration.

The specific procedures of this experiment are as follows:

- 24 students are averagely divided into two groups according to their general translation performance to ensure close translation competency;
- The first group of students (No.1 to No.12) translate text 1; the second group of

students (No.13 to No. 24) translate text 2;

- A brief introduction to the military translation parallel corpus is made; limited by time and condition, only some simple and basic functions are introduced;
- Then, the second group of students translate text 1 and the first group of students translate text 2 by using the corpus;
- At the end of the experiment, the students' comments of the corpus are collected.

4. Result Analysis

4.1 Translation Speed

As is considered above, both translation tests are required to be finished in one and half hours. The first test is a little bit prolonged, and many students cannot hand in their translation on time, which is directly caused by their translation competence and the textual difficulty degree.

The second test is much better. All subjects hand in their translation in 70 minutes. Although there is no big difference between the durations of the two tests, the translation speed is much improved if we consider the fact that this is the first time that the subjects operate a corpus and they are not familiar enough with the procedures yet.

4.2 Translation Quality

4.2.1 Examples

This paper will list three examples which are used to illustrate the roles the military

translation parallel corpus plays in promoting the students' competences of foreign language, native language and subject-field knowledge[④].

The first example is mainly to show that the corpus plays a positive role in promoting the subjects' ability of foreign language understanding. There is such an expression in the test:

Example 1:

… as we continue to apply ***emerging*** information technologies…

The recommended Chinese translation of this part is:

……既然我们要继续应用**新的**信息技术……(Chinese pinyin of these characters are: *jiran women yao jixu yingyong xinde xinxi jishu*)

The word "emerging" is paraphrased in *Oxford Advanced Learner's English-Chinese Dictionary* as "显现(*xianxian*)，出现(*chuxian*)，涌现(*yongxian*)，显露(*xianlu*)", but some students translate the word as "紧急(*jinji*)，急需(*jixu*)". In the corpus, 77.37% of "emerging" is translated as "新兴的(*xinxingde*)". The first group of subjects obviously hasn't understood "emerging" properly and their translation of this word is miscellaneous and inexact. Aided by the corpus, the second group of subjects gets a lot of improvement and the rate of correctly translating the word significantly increases as is shown in the following table:

[④] It's easy to understand subject-field knowledge. As for foreign language competence and native language competence, the author of this paper understands this way in the result analysis: generally speaking, foreign language competence refers to the process of translators' understanding the source text and native language competence the process of producing the target text in native language.

Table 1. A comparison of non-corpus-based and corpus-based translations of *emerging* in Text 1

Group 1	－Emerging[5]		Group 2	+Emerging	
1	新兴的	√	13	新的	√
2	逐渐成形的	×	14	涌现出的	×
3	形成	×	15	现有	×
4	不断涌现的	×	16	显现出来的	×
5	初现的	×	17	新的	√
6	现有	×	18	* [6]	×
7	逐渐显露的	×	19	新显现的	√
8	新的	√	20	急需的	×
9	紧急	×	21	新出现的	√
10	新发现	×	22	*	×
11	形成中的	×	23	新兴	√
12	不断涌现的	×	24	新兴的	√
Correct translations		2	Correct translations		6

The comparison may tell us that although traditional dictionaries have tried their best to provide language learners with explanations and sample sentences, they are still very insufficient in terms of register of the words and real context of language use. As a result, the translators quite often cannot get satisfying references from traditional dictionaries only.

[5] "-emerging" means non-corpus-based translations of "emerging"; "+emerging" means corpus-based translations of "emerging". In the following tables, the same rule is used.

[6] The symbol * is to indicate that the word "emerging" is not translated.

The second example is to illustrate the corpus's positive influence on the subjects' subject-field knowledge.

Example 2:

Network Centric Warfare （*NCW*） → 网络中心战 (*wangluo zhongxinzhan*)

Platform centric warfare （*PCW*） → 平台中心战 (*pingtai zhongxinzhan*)

The above are two central terms in the test whose proper translation cannot be found in ordinary dictionaries, so almost nobody in the subjects in Group 1 has translated the two terms correctly. Actually they will be very easy to be translated if the corpus is referred to, so almost all the subjects in Group 2 have translated correctly. Here two points need to be explained further: one is that in Group 1 three subjects have translated NCW correctly; we guess this may be because they have ever read related documents; the other is that in the corpus there is no such a term as PCW, but there are "platform" and "platform centric" which are respectively translated as "平台 (*pingtai*)" and "平台中心(*pingtai zhongxin*)", so easily the subjects can infer that PCW can be translated as "平台中心战 (*pingtai zhongxinzhan*)".

Table 2. A comparison of non-corpus-based and corpus-based translations of *NCW* and *PCW*

Group 1	－NCW	－PCW	Group 2	+NCW	+PCW
1	×	×	13	√	√
2	√	×	14	√	√

3	×	×	15	√	×
4	×	×	16	√	×
5	×	×	17	√	√
6	×	×	18	√	×
7	×	×	19	√	√
8	√	×	20	√	√
9	×	×	21	√	√
10	×	×	22	√	√
11	√	×	23	√	√
12	×	×	24	√	√
Correct translations	3	0	Correct translations	12	9

The third example is to show that the corpus also plays an important role in promoting the subjects' native language ability.

Example 3:

The ***commercial*** experience has shown how information can substitute for material…

(Emphasized by the author of this paper)

The acceptable translation should be "从**民用领域**的演变中可以了解到信息是怎样代替物质……" (Chinese pinyin: *cong minyong lingyu de yanbianzhong keyi liaojiedao xinxi shi zenyang daiti wuzhi*)

The translations of the word "commercial" of the two groups are as follows:

Table 3. A comparison of non-corpus-based and corpus-based translations of *commercial*

Group 2	－commercial		Group 1	+commercial	
13	商业		1	商业	
14	商业的		2	商业	
15	经济上的	×	3	商业	
16	商业		4	商业的	
17	商场的	×	5	商业上的	
18	商场的	×	6	民用上的	√
19	商业上		7	商业上的	
20	商业		8	商业的	
21	商场上的	×	9	民用	√
22	商业领域		10	商业	
23	商业经验		11	商业的	
24	商业经验		12	民用	√

Actually both groups have understood the meaning of “commercial experience”, but most of them have difficulty in expressing the meaning in proper Chinese. Generally speaking, “军用” (*junyong* meaning “military”) and “民用” (*minyong* meaning “civilian”) are an opposite pair in Chinese and in this context “commercial” is very well translated as “民用”. Although the translation “商业” (*shangye* literarily meaning “commercial”) is inexact, it’s almost acceptable; but the translations “商场” (*shangchang* meaning “department store”, “经济” (*jingji* meaning “economy”) are too farfetched. Most of the first group of subjects translate “commercial” as “商业” under

the aid of the corpus and three of them translate the word exactly as “民用”. As a contrast, the second group doesn’t translate the word properly; only four subjects translate it as “商业” which is not so acceptable in the context.

4.2.2 General Trend Observation

In this part the mean values of the non-corpus-based and corpus-based translation mistakes are to be analyzed to try to find out the general trend and regularity.

Table 4 shows the total numbers of three kinds of mistakes that the subjects have made in their translations of text 1. The left side lists the numbers of mistakes made by group 1 without the aid of the corpus; the right side lists the numbers of mistakes made by group 2 with the aid of the corpus. In the second and the fourth columns the three numbers from left to right represents respectively **foreign language mistakes**, **subject-field knowledge mistakes** and **native language mistakes**.

Table 4. A comparison between the numbers of translation mistakes made by group 1 in translating text 1 without aid of the corpus and those of group 2 with aid of the corpus

Group 1	Non-corpus-based	Group 2	Corpus-based
1	2—10—2	13	6—7—1
2	5—6—7	14	7—8—3
3	13—7—4	15	4—5—3
4	9—6—1	16	2—6—2
5	4—7—9	17	3—6—4

6	2—10—2	18	5—8—2
7	3—11—1	19	3—5—0
8	4—12—1	20	4—7—1
9	7—9—3	21	9—9—2
10	3—13—0	22	5—7—2
11	4—10—1	23	2—6—3
12	10—8—0	24	3—9—0
Total number of all mistakes	66—109—31	Total number of all mistakes	53—83—23
Average of each kind of mistakes	5.5—9.08—2.58	Average of each kind of mistakes	4.42—6.92—1.92

Table 5 shows respectively the numbers of the three kinds of mistakes made by group 1 without aid of the corpus and those made by group 2 with aid of the corpus. As in table 4, the numbers in the second and the fourth column represent respectively foreign language mistakes, subject-field knowledge mistakes and native language mistakes.

Table 5. A comparison between the numbers of mistakes made by group 2 in translating text 2 without aid of the corpus and those of group 1 with aid of the corpus

Group 2	Non-corpus-based	Group 1	Corpus-based
13	3—11—6	1	4—6—4
14	5—11—4	2	2—9—2

15	3—8—7	3	10—5—4
16	7—7—5	4	7—7—3
17	9—9—2	5	4—6—3
18	7—7—2	6	5—8—1
19	3—6—5	7	2—6—2
20	7—7—3	8	2—7—1
21	9—9—2	9	6—9—0
22	4—10—0	10	4—7—2
23	6—14—1	11	2—9—2
24	7—12—2	12	7—7—1
Total number of all mistakes	70—111—39	Total number of all mistakes	55—86—25
Average of each kind of mistakes	5.83—9.25—3.25	Average of each kind of mistakes	4.58—7.17—2.08

We can draw the following conclusions through analyzing the data in table 4 and table 5.

(1) Text 2 is slightly more difficult than text 1. This is self-evident when we compare the average of mistakes in the translations of text 1 (Table 1) and those in text 2 (Table 2). As is shown in Table 1 and Table 2, no matter with or without the aid of the corpus, the average of each of the three kinds of translation mistakes (i.e. foreign language mistakes, subject-field mistakes and native language mistakes) in text 1 is smaller than that in text 2.

(2) The corpus has significant positive influences on the subjects' translation

competences (productions). The data in Table 4 and Table 5 show clearly that no matter in which text, no matter what kind of mistake, no matter the total number or the average of mistakes, the subjects' translations with aid of the corpus have obviously fewer mistakes than those without aid of the corpus.

4.2.3 General Trend Shown by SPSS[⑦]

After the simple analysis of the average values of the data, we are going to investigate the experiment results from the perspective of statistics.

Table 6. A comparison between the total numbers of each subject's mistakes in non-corpus-based and corpus-based translations

Group 1	1	2	3	4	5	6	7	8	9	10	11	12
Total number of non-corpus-based mistakes	14	18	24	16	20	14	15	17	19	16	15	18
Total number of corpus-based mistakes	14	13	19	17	13	14	10	10	15	13	13	15
Group 2	13	14	15	16	17	18	19	20	21	22	23	24
Total number of non-corpus-based mistakes	20	20	18	19	20	16	14	17	20	14	21	21
Total number of corpus-based mistakes	14	18	12	10	13	15	8	12	20	14	11	12

⑦ SPSS: The acronym of Statistical Package for Social Sciences

The analyzing tool used here is Statistical Package for Social Sciences (SPSS). There are mainly two analysis items: one item is to carry out respectively an independent sample *t*-test for the three kinds of translation mistakes in Table 4 and Table 5; the other is to carry out a paired sample *t*-test for the total numbers of the subjects' mistakes in the two tests (both non-corpus-based and corpus-based) in Table 6. The *t*-test results are as follows:

(1) In the independent sample *t*-test of the three kinds of translation mistakes, only the subject-field understanding *t*-test reveals itself a significant difference ($p1 = 0.011 < 0.05$; $p2 = 0.016 < 0.05$)[⑧]. This shows that the experiment is sensitive to the students' subject-field translation competence. In other words, the use of the military translation parallel corpus in translation can dramatically reduce mistakes in subject-field understanding and expressing and the quality of the translation is significantly improved.

(2) The paired sample *t*-test of the subjects' total mistakes in the two translation tests shows a high degree of significant difference ($p3 = 0 < 0.01$)[⑨]. This means that the military translation parallel corpus plays a very important role in improving the overall quality of the subjects' translations.

[⑧] p1 stands for the value of significant difference between the mistakes made in corpus-based and non-corpus-based translations of text1 described in table 4; p2 is the value of significance difference of that of text 2 in table 5.

[⑨] P3 is the value of significant difference between the overall mistakes made in translations listed in table 6.

4.3 The subjects' comments

18 of the 24 subjects write down their comments after the second translation test. Almost all participants think that the corpus is very helpful for their translation practice. No.7, No.8, No.12 and No.15 subjects think that the corpus "can provide detailed military background knowledge which helps them understand the source text better"; No.6, No.8 and No.24 subjects think that the corpus "can provide authoritative translation reference which helps improve the exactness of the translation"; No.12, No.14 and No.22 subjects think that the corpus is particularly helpful for terminology translation; No. 6 subject thinks that his translation speed is obviously improved with the aid of the corpus.

5. Conclusion

The experiment shows that the military translation parallel corpus can significantly improve the translation speed, subject-field knowledge and overall translation competence of the subjects. The corpus even plays a very positive influence on the subjects' foreign language competence and native language competence. With the enlargement of the corpus and more ease of operating the corpus, this kind of specialized military translation parallel corpus can play more significant roles in military translator training.

References

Bowker, Lynne (1998), "Using Specialized Monolingual Native-Language Corpora as a

Translation Resource". *Meta* 43(4): 631-651

Liang, Xiaobo et al (2008), "Application of Military Corpora under Information Condition". National Defense Science and Technology (2): 51-57.

Liu, Qun (2000), "On the Quality Standards and Knowledge Structure of Science and Technology Translation". Chinese Science and Technology Translators Journal (1): 37-39.

May Fan & Xu Xunfeng (2002), "An Evaluation of an Online Bilingual Corpus for the Self-learning of Legal English". Available online, http://www.elsevier.com/locate/system. 2002

Sykes, J.B (1989), *The Translator's Handbook (2nd Ed.)*. London: Aslib, pp35-39.

Wei, Yaoyu & Lei Lei (2006), "The Aid of Parallel Corpora for Chinese English Learners' Translation Practice", Foreign Language Teaching Elementary Education (4): 28-30.

Yang, Huizhong (2002), *An Introduction to Corpus Linguistics*, Shanghai: Shanghai Foreign Language Education Publishing House, p29.