

Application of DIY Corpora to Translator Education: A Practice-based Study

Hongmei Han*

School of Foreign Languages
Institute of Disaster Prevention
Sanhe, China
hanhongmei@cidp.edu.cn

Enqing Li

School of Foreign Languages
Institute of Disaster Prevention
Sanhe, China
lienqing79@163.com

Abstract—With the development of big data technology, the application of corpora to translation studies and translator education has drawn the attention of more scholars and instructors. Although translation studies by way of DIY (“Do-It-Yourself”) corpora have been carried out, their application to translator education requires more exploration. This study demonstrates that DIY bilingual translation corpora and monolingual comparable corpora targeted at translator education are of great value in promoting the effects of translator education, making translation students sensitive to the differences between Chinese and English and raising their cross-cultural communication consciousness. The study proves that the application of corpora can be a feasible approach to translator education.

Keywords—DIY corpora, translation corpora, comparable corpora, translator education

I. INTRODUCTION

Research on translator education is an important branch in corpus-based translation studies. Translation students can be motivated to actively participate in the learning process through collecting and evaluating texts, extracting terminology, and establishing correspondence between two languages. [1] Besides, “Parallel corpora and comparable corpora have complementary roles to play in the translator training environment” [2]. Therefore, the use of corpora in translator education has broad prospects. However, the application of corpora to Chinese-English translator education remains at the discussion stage. Based on the studies carried out by predecessors and peers [3-7], this research will “walk the talk” and explore the use of DIY corpora in translator education.

II. CORPUS CONSTRUCTION

The DIY translator education corpus is different from those large-scale general corpora in that its design follows the practice of translator training. It covers a wide range of subjects and genres, which meets the variety requirement for translation teaching materials. Based on the text type division of Mary Snell-Hornby[8], the current study extends the general text types to more specific text types. Three categories were designed for the translator education corpus, namely the literary classics translation sub-corpora, the general texts translation sub-corpora and the professional documents translation sub-corpora. In each category, sub-corpora were

constructed according to the needs of translator education. The framework of the translator education corpus is shown in Fig. 1.

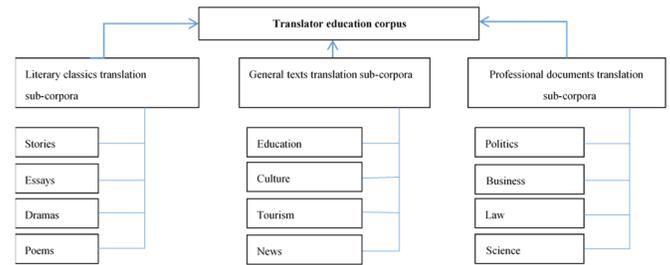


Fig. 1. The framework of the translator education corpus

The sub-corpora in the translator education corpus were required to be teaching-targeted and each contains an English-Chinese sub-corpus and a Chinese-English sub-corpus. After the completion of the sub-corpora, the teaching practice would be carried out.

In addition, the role of monolingual corpora in translator education would not be ignored. Therefore, besides constructing sentence-aligned bilingual parallel sub-corpora for each category, corresponding monolingual sub-corpora were designed according to needs.

This study takes one of the subcorpora of the translator education corpus, a Chinese-English education-related sub-corpus — University Profile Corpus (hereinafter referred to as UPC) — as an example to explore the application of DIY corpora to translator education.

III. A CASE STUDY OF DIY CORPORA APPLICATION TO TRANSLATOR EDUCATION

UPC includes a Chinese-English translation corpus of Chinese university profiles (hereinafter referred to as TC) and a monolingual comparative corpus of the university profiles from the countries with English as their official language (hereinafter referred to as CC). CC comprises two sub-corpora — the English translations of Chinese university profiles (hereinafter referred to as ET) and 100 English university profiles from the countries where English is an official language (hereinafter referred to as EO). Both TC and CC play a role in the exploration of translation strategies and methods and contrastive linguistic studies.

Supported by Research Foundation for University Teachers, China Earthquake Administration(No. 20150120), Research and Practice Project of Hebei Higher Education Teaching Reform (No. 2016GJJG249), Fundamental Research Fund for the Central Universities(No. ZY20180222) and Translator education Team Construction Fund(No. JT201605)

A. Translation Strategies and Methods Exploration

The content of TC is made up of profiles of the top 200 Chinese universities in 2016 (Wu Shulian Ranking) downloaded from their official websites before September 1, 2017. The overall strength of the universities and the authority of the official websites guarantee the reliability of the corpora, which ensures their applicability to translator education.

Applying the corpus to translator education started early at the initial stage of the corpus construction. The instructor assigned homework first. Each student was to collect both Chinese and English texts from assigned universities, save the aligned texts at paragraph level in an excel file if possible, and then write the translation strategy they observed. Through participation of the corpus construction, the students found out the complexity of Chinese-English translation through observation and comparison, because the translations of the Chinese university profiles presented various strategies—full translation, partial translation, trans-editing and rewriting. Now they had a first taste of the translation strategies adopted by professionals concerning profile translation. The texts submitted by the students needed checking and were in separate files, so the instructor had to further process the texts to complete the corpus. The final corpus presented by the instructor was surprising: only 29 of the 200 Chinese university profiles matched fully or partially with their English translations, while 155 university profiles were largely trans-edited or rewritten in English. Besides, there were 16 universities whose English profiles did not exist or could not be downloaded. What made things complicated was that some universities failed to update the English version after the Chinese version was updated, resulting in inconsistency of some numbers. However, this did not affect the study of languages, so such university profiles were also put into the bilingual parallel corpus. In order not to affect understanding, the inconsistent numbers were modified during the processing of the sentence alignment. The unexpected proportion of translated university profiles shows the advantage of the corpus application to translator education. It objectively presented a translation phenomenon in reality with data, that is, trans-editing or rewriting the profiles was the strategy choice of most translators.

The corpus construction completed, the instructor began to apply the corpus to classroom teaching. The biggest obstacle in translating the Chinese profiles from the students' point of view was the Chinese culture-specific expressions. So the university profile sub-corpus TC was first used to help with the translation of such expressions. The instructor used the translation corpus search tool BFSU ParaConc 1.2.1 to upload the corpus. And next, the search terms were determined. The Chinese culture-specific expressions are widely used in the university profile corpus, and many of them are in quotation marks. Therefore, “*” was used as the search item to locate these expressions. Altogether 201 Chinese-English translation sentences were retrieved, which were presented along with their sentence contexts. Next, the instructor encouraged the students to do an in-depth study of all the search results. And then a discussion was carried out. The instructor studied the results in advance when preparing lessons so that the tasks

could be arranged in a targeted manner. In truth, the research process was presented to the students during the discussion.

Proper nouns about projects can be taken as an example to illustrate how to use corpora to assist translator education. The corpus statistical results showed that higher education-related projects involve university development, talents selection, personnel training and others. All the instructor needed to present to the students were the bilingual projects that had been retrieved. As the instructor presented the retrieved corpus items, the students were asked to observe and summarize their findings about the translation methods. Their findings can be summarized as follows: a. Quite a few methods were adopted to translate proper nouns about education projects, especially the national-level ones; b. Literal translation was a dominant method for the translation of the proper nouns about education projects, but other methods were also used such as omission, addition, liberal translation, combination of literal translation and liberal translation.

The findings were the results of observation. And a further discussion on the results was held. For example, what translation methods are better than the others and can be recommended for translating the proper nouns about education projects? Although there is no universally applicable method in translation, should there be a unified translation for national projects? During the discussion, the students actively expressed their ideas. The instructor simultaneously displayed relevant contexts as the discussion was going on. Through interaction, the students advanced their understanding of the translation strategies and methods used for those proper nouns.

B. Translation Reference and Verification

Comparing the English translation corpus and the original English corpus on the same topic enables translation students to deepen their understanding of the two languages and the differences between English and Chinese in expressing similar ideas, thus arousing the students' cross-cultural awareness and improving their translation. The following is a comparison study between ET and EO. Through this study, similarities and differences between English translation and original English in expressing similar content can be discovered.

In the construction of the Chinese University Profile Corpus, the English translation of the 29 university files and the English trans-editing or rewriting texts of the 155 university profiles were stored in one folder as ET. And EO was sourced from the profiles of 2017 QS top 100 world universities whose official language is English.

First of all, Wordsmith7.0 was used to get the general statistics of both ET and EO, such as word types, tokens, type/token ratio, standard type/token ratio, mean word length and mean sentence length, which sketched the images of ET and EO. In this way, the students had a macro understanding of the English translation of Chinese university profiles (including translation, trans-editing and rewriting) and the original English university profiles like that the vocabulary in the translated texts is not as rich and varied as that in the original English texts; and there are more long words and long sentences in the translation, which makes the texts more difficult. The findings led to more discussions on the reasons.

And the reasons analyses inspired the students to pay special attention to the diction and sentence structures in their own translation practice.

In terms of content, word frequency displayed the respective focus of ET and EO. Table I shows the top 20

content words in ET and EO (A longer list can be shown if necessary). The instructor presented the data and the students observed.

TABLE I. THE TOP 20 CONTENT WORDS IN ET AND EO

ET					EO				
Word	Freq.	%	Texts	%	Word	Freq.	%	Texts	%
UNIVERSITY	2735	1.98	184	100.00	UNIVERSITY	1671	1.38	99	99.00
NATIONAL	1344	0.97	153	83.15	RESEARCH	1026	0.85	98	98.00
EDUCATION	1207	0.88	176	95.65	STUDENTS	772	0.64	97	97.00
RESEARCH	1087	0.79	170	92.39	WORLD	664	0.55	94	94.00
SCIENCE	787	0.57	153	83.15	FIRST	418	0.34	76	76.00
ENGINEERING	755	0.55	153	83.15	NEW	414	0.34	78	78.00
STUDENTS	746	0.54	165	89.67	COLLEGE	379	0.31	67	67.00
CHINA	707	0.51	158	85.87	CAMPUS	372	0.31	75	75.00
KEY	680	0.49	151	82.07	EDUCATION	353	0.29	90	90.00
SCHOOL	643	0.47	121	65.76	SCHOOL	344	0.28	60	60.00
LEVEL	622	0.45	129	70.11	MORE	337	0.28	75	75.00
UNIVERSITIES	597	0.43	162	88.04	UNIVERSITIES	280	0.23	76	76.00
PROGRAMS	584	0.42	140	76.09	FACULTY	260	0.21	69	69.00
TECHNOLOGY	553	0.40	123	66.85	TEACHING	257	0.21	67	67.00
FIRST	527	0.38	132	71.74	COMMUNITY	251	0.21	68	68.00
INTERNATIONAL	523	0.38	144	78.26	ACADEMIC	250	0.21	83	83.00
DISCIPLINES	517	0.37	132	71.74	STUDENT	243	0.20	77	77.00
CHINESE	510	0.37	142	77.17	GLOBAL	240	0.20	63	63.00
PROVINCIAL	503	0.36	109	59.24	INTERNATIONAL	239	0.20	69	69.00
TEACHING	497	0.36	131	71.20	SCIENCE	232	0.19	64	64.00

The findings were that some of the top 20 high-frequency words overlapped in the two corpora. These overlapped high-frequency words validated the common theme of the two corpora: the main tasks of universities are education and scientific research; all universities try to promote themselves (FIRST) and going international is an important strategy (INTERNATIONAL). The different high-frequency words between the two corpora were left for the students to explore. The information represented by these words, especially the information focus in EO, provided reference for the students when they translate profile type texts, especially when they adopt the rewriting strategy.

The application of high frequency word list provides important reference for target readers-oriented rewriting. Besides, the monolingual corpus, such as EO in this study, has another function: to verify whether certain translations are acceptable for native English readers. In the teaching practice, the 16 Chinese university profiles without English translations became translation assignments for the students. In the translating process, the students could use not only TC to find translation strategies, translation methods and translations of corresponding expressions, but also EO to verify their own translations or find a reasonable language structure. The translation of the term “shiziduiwu” can be taken as an example.

“Shiziduiwu” is a frequently mentioned term in the university profiles. The corresponding translations the students

could find in dictionaries and ET were “faculty”, “faculties” or “teaching staff”. Then, what is the situation of the three expressions in EO? What words collocate with them? Searching them in EO through AntConc3.4.3 showed the differences of the co-occurring words. “Faculty” can be the most convincing equivalent to “shiziduiwu”, for it is used most frequently in this sense (up to 260 concordance lines) and usually co-occurs with such words as numbers, staff, and students. The plural form “faculties” is mainly used to mean “colleges” or “departments”. “Teaching staff” is used at a lower frequency — seven concordance lines in total, and appears in specific situations rather than macro-introduction contexts, such as “the University’s academic teaching staff” and “95 students and 4 teaching staff”. Observation of the differences between the three expressions enabled the students to choose the correct translation. In the meantime, the contexts in which these expressions appear provide a grammatical structure reference for the translation of the Chinese sentences in which the term “shiziduiwu” appears.

In summary, linguistic comparison between ET and EO enables translation students to be aware of linguistic differences between translated English and original English and assists them to cross the linguistic and cultural barriers and make their translation a natural representation in English.

IV. CONCLUSIONS

The current study, based on the achievements of corpus linguistics and translation studies, explored the application of

DIY corpora to translator education through a case study—the application of Chinese-English translation sub-corpus — UPC — to translator training. The study found that translation students’ participation in the process of corpus creation can deepen their understanding of the differences between Chinese and English, and assist them to understand the translation strategies, thus guiding them in their translation strategy choice. In addition, translation corpora can be used to discover how professional translators deal with translation difficulties (such as the translation of culture-specific expressions) and monolingual corpora can be applied to checking the usage of some expressions in the translated language, thus verifying whether the translation conforms to the target language rules. As it turned out, corpora can be effective in improving the students’ translation.

The creation of each sub-corpus in the translator education corpus was a complicated process. Although all sub-corpora have something in common, this process was full of unexpected changes due to the differences in style and research target. In constructing UPC, we experienced a huge contrast between expectations and reality, but this is precisely the charm of applying corpora to translator education. The reality of language and objective data may go beyond our cognition and provide more possibilities for translator education.

ACKNOWLEDGMENTS

The software used in this study includes BFSU ParaConc1.2.1 developed by Xu Jiajin, Liang Maocheng & Jia Yunlong, WordSmith7.0 developed by Mike Scott and AntConc3.4.3 developed by Laurence Anthony. Our gratitude goes to all these programmers.

REFERENCES

- [1] Hu K.B. & Mao P. F., “Corpus translation studies abroad: a critical review,” *Contemporary Linguistics*, no. 4, pp. 388, 2012.
- [2] Pearson, Jennifer, “Using Parallel Texts in the Translator Training Environment”, in F. Zanetti, S. Bernardini, D. Stewart. 2007. *Corpora in translator education*. Beijing: Foreign Language Teaching and Research Press, pp. 15-16, 2007.
- [3] Zanettin, F., “Bilingual Comparable Corpora and the Training of Translators,” *Meta*, no. 4, pp. 616–630, 1998.
- [4] Bowker, Lynne, “Towards a Methodology for a Corpus-Based Approach to Translation Evaluation,” *Meta*, vol. 46, no. 2, pp. 345-364, 2001.
- [5] Bernardini, S, Stewart, D & Zanettin, F., “Corpora in Translator Education: An Introduction,” In F. Zanetti, S. Bernardini, D. Stewart. 2007. *Corpora in translator education*. Beijing: Foreign Language Teaching and Research Press, pp. 1-14, 2007.
- [6] Wang, K. F., “The use of parallel corpora in translator training,” *Media in Foreign Language Instruction*, no. 6, pp. 27-32, 2004.
- [7] Liu, Z.Q., & Liu, D., “Application of multi-media computer technology and corpus-based approach in the renovation of the translation course,” *Foreign Languages & Their Teaching*, no. 8, pp. 27-30, 2009.
- [8] Mary Snell-Hornby, “Translation Studies: An integrated approach,” *Foreign Language Teaching and Research Press*, p. 34, 2006.