

A Corpus-based Translation Class

YAO Xing-an (姚兴安)

Business School of Xi'an International Studies University Xi'an, Shaanxi, 712000

Abstract: The past decades have seen the adoption of the corpus-based approach to translation studies. Corpus-based Translation Studies (CTS) is now recognized as a major paradigm informing a wide array of studies in the discipline. The present paper tries to show that online English and Chinese corpora are very useful tools for translation teaching and learning. It is intended to introduce the application of them in class activities. It concludes that a corpus is an available channel for obtaining abundant and illustrative translation teaching and learning materials and that the corpus-based approach has shown great potential in translation studies and thrown a light into a brand new approach to improving students' translation competence.

Keywords: corpus-based approach; translation teaching; translation competence

0. Research Rationale

Translation studies as a whole has witnessed great developments in recent years. Among them, one of the most prominent trends is towards Descriptive Translation Studies (DTS). Under this paradigm, translation is no longer viewed as a static process of text transfer from the source language to the target language, but as a dynamic process combining many other factors, such as text purpose, sponsorship, readership, etc. However, in the field of translation teaching, or

translator education, the conventional approach still dominates the field. The weaknesses of this methodology are obvious and therefore other methodologies need to be explored.

Recently, the Corpus-based Translation Studies (CTS) has been gaining popularity in this arena. Along with the fast developments of corpus linguistics, concordance tools and software also flourish and are constantly being improved. In the area of translation teaching or translator education, corpus begins to show its values and advantages. Conventionally, translator education is restricted to the pattern of teaching fragmented sentences. In most cases, it is always conducted in such a manner that the teacher is the only lecturer who does all the talking while the students are considered as passive recipients. The fact is that not only is the teaching method boring, but the translation textbooks are also monotonous and outdated. The use of corpus can supplement the deficiencies of conventional teaching methods.

1. Research Structure

This research involves an empirical paradigm and Internet-related resources to explore the feasibility and usefulness of corpus-based translation teaching. In addition, the online corpora built by other researchers are also used as a supplement in the translation experiment of the current research.

The empirical research is conducted with a pilot study which is based on China Translation Dictionary (<http://www.chinafanyi.com>). The participants of this pilot study are third year students. The subjects in this experiment are divided into an experimental group and a control group with the former completing the translation of an assigned text using corpus resources and the latter using conventional resources. A qualitative analysis is then carried out to compare the translated texts of the two of corpus-assisted translation.

2. Research Aims & Methods

The author tried to posit answers to the following questions:

(1) Is a web-accessible corpus a useful resource for translation trainees?

(2) Is the corpus a useful tool and enable students to overcome the limitations of conventional reference tools such as dictionaries?

The main theoretical model of the study is a comparative one, which examines how two different groups of student translators of similar language competence can perform on two different situations, one using the conventional resources while the other using the extra resource of computerized corpus. The current research is empirical in nature, but the research also includes qualitative and conceptual research because the research topic is about translation teaching and students' translation competence.

3. Previous Research in CTS

With the rapid development of computer technology since the 1980s, more achievements have been made in the field of corpus linguistics. A number of large-scale corpora were completed in this period. It is generally acknowledged that Mona Baker (1993) is the first scholar to propose the use of corpus to investigate the phenomena of translation. Baker (1993:243) predicted that the availability of large corpora of both original and translated texts, together with the development of a corpus-driven methodology, would enable translation scholars to uncover "the nature of translated text as a mediated communicative event." Since then, corpus has been widely used as a research tool by many scholars in the translation field. Some researchers even try to utilize corpus to help student translators solve problems encountered in the process of learning translation.

In the area of corpus-based research, the Center for Translation Studies in the Department of Language Engineering, UMIST, led by Mona Baker, has set up the first comparable corpus in the world in 1995. The corpus is named as Translational English Corpus (TEC). Today, the corpus is still being extended to include more texts. The corpus not only collects texts of modern English, but also English texts translated from other languages. In addition, the corpus includes texts of Non-European languages, such as Japanese and Chinese.

The corpus-based research in China started much later than that of the western world. Corpus-based studies of Chinese translated texts are relatively less than those of European languages. Xiao & McEnery (2004) study aspect markers of Chinese translated texts in Lancaster University and Chen (2003) deals with explicit use of more causal conjunctions are among the distinctive ones. Beijing Foreign Studies University is a pioneer in the research of this area and scholars there have set up the largest Chinese-English Parallel Corpus in the world (Wang Kefei 2003, 2004).

4. A Pilot Study

This pilot study consists of a translation experiment that was carried out at Xi'an International Studies University in 2007. The experiment uses China Translation Dictionary and several other search engines including "Google" as reference corpora in our experiment.

Hypotheses

1) It is hypothesized that translators who translate with the help of a web-resource will perform better than those using conventional resources though they are of the same language ability.

2) It is hypothesized that students will make great progress in their translation performance as well as translation competence by receiving training in a corpus-based manner.

Participants

In this experiment, the subjects are 30 third year students from Business School of Xi'an International Studies University. It is believed these students would be relatively candidates for the experiment for the following reasons: all the participants are third year undergraduate students and their language ability is acceptable for carrying out the task; the participants in the experimental group are familiar with basic computer skills and have some years of Internet experience.

Translation Text in the Experiment

A text from China Daily is chosen for the translation experiment. The topic of this translation text is about Qomolangma. The text contains 414 words, which is not very long and students can finish it within a period of time. Moreover, since the text is about general knowledge, the current researcher feels competent to evaluate their work. The source text used in the experiment can be found in Appendix.

Experiment

For the purpose of this experiment, all students were translating out of their native language (Chinese) into their foreign language (English). The 30 students are divided into two groups: experimental group and control group. Participants of the control group would use the conventional resources while the experimental group would use both conventional and corpus resources. It is worth noting that the issues relating to the students' foreign-language skills were beyond the scope of this study. Therefore, one important criterion for this is to ensure that both groups have relatively the same English level. To ensure that, the TEM4 exam scores of the student participants are taken as one benchmark to divide the two groups (Table I).

Table I

	TEM-4 Scores														
Experimental Group	88	86	86	85	85	84	83	81	80	79	76	76	75	72	68
Control Group	89	87	87	85	84	82	82	80	80	80	78	77	72	69	66

The resources used by the two groups of students are also different. Three online corpora are used in the experiment. Besides, the web search engine Google is treated here as a big corpus (Table II).

Table II

Resources	Monolingual Dictionaries
Accessible to Both Groups	Bilingual Dictionaries
Resources	Electronic Dictionaries
Accessible Only to Experimental Group	<p>Babel Corpus Online (http://bowland-files.lancs.ac.uk/corplang/babel/babel.htm)</p> <p>China Translation Dictionary (http://www.chinafanyi.com)</p> <p>Thesaurus Website (http://www.thesaurus.com)</p> <p>Google Search Engine (http://www.google.com)</p>

The researcher would like to select the following sentence “珠穆朗玛峰四周地形极其险峻，环境异常复杂” for an example to illustrate the choice of words. Most of the students translated “环境” into “environment”, which, in fact, is inappropriate. “Environment” refers to

“the air, water, and land in which people, animals, and plants live” (Longman Dictionary of Contemporary English). It does not reflect the real sense of the source text. The word “topography” should be used to refer to “the shape of an area of land, including its hills, valleys, etc”. The following are more collocations related to 环境 in the corpus of China Translation Dictionary:

○生活状况、居住环境、生活环境、生活条件living condition

○空气质量 环境空气质量评定用分层采样法Air quality; stratified sampling method for assessment of ambient air quality

○年轻人适应新环境的能力是他们的一种优点。

The adaptability of youth to new surroundings is one of their good qualities.

○将商业利益置于环境保护之上、重视商业利益高过环境保护
puts trade interests above green issues

○加强教育和管理，落实责任制，创造良好的社会治安环境。

We should intensify education and management to reinforce the responsibility system and bring about a sound environment for public security .

○环境的确艰难，但港人应和政府团结一致，以积极思想共创未来。

Yes, things are bad, but it's up to the community working together with the government to think about the positives and to think about how we can create a future for ourselves.

○【谚】环境不同人也变，人的行为因时因地而异；此一时彼一时。

Circumstances alter cases.

○开放的××以一流的政府服务和一流的投资环境，热忱欢迎国内外客商来××投资置

业

With first-class government services and a top-rank investment climate, the opening ×× readily welcomes domestic and foreign businessmen to make investments

○我们看见各种的动物，它们生活的环境都建造得尽可能接近它们的原产地。我们看见狮子，老虎，大象，美丽的鸟类，以及猴子等等。除此之外，我们能够欣赏有趣的猴子秀。

We saw all kinds of animals in dwellings that were as close as possible to their original habitats. We saw lions, tigers, elephants, beautiful birds, monkeys and so on. Besides these, we were able to see an interesting monkey show.

○在长期执政的条件下，在对外开放和发展社会主义市场经济的环境中，党必须十分注重防范各种腐朽思想的侵蚀，维护党的队伍的纯洁。

"Against the Background of its long-term governance and China's opening up and development of the socialist market economy, the Party must be on full alert against corrosion by all decadent ideas and maintain the purity of its membership."

○在使用铅的地方工作和暴露于农药环境下也是危险的因素。1996年美国完成了含铅汽油的逐步淘汰，世界各地也正在实施类似的禁令。

Working where lead is used and exposure to some insecticides are other risk factors. The U.S. phaseout of lead in gasoline was completed in 1996; similar bans are being implemented worldwide

○深信我们大家都同样爱护鸟类以及关心牠们的生态环境，你们的支持是推动鸟类和生态环境保育不可缺少的动力。

You are among our many friends who care about birds and their habitat. Your support is essential, helping us to meet very demanding requirement in aiding the conservation of birds and their habitat.

In fact, these examples might motivate students to come up with their own choice of translations. Corpus users show more confidence when rendering translations to certain words. However, dictionary users are always worried that they might not be correct. This is typically seen in the translation of this sentence by dictionary users.

Another example is also found here in the translations between dictionary users and corpus users. The corpus users also show more varieties in their rendering of translated texts: 当时中国科学家测定珠峰高度为8848.13米。这一数据一直作为中国对珠峰高度的采用值沿用至2005年。

Version 1: Chinese scientists then gauged the height of it as 8848.13 meters. This data had been adopted by Chinese government till 2005.

Version 2: The team measured its height as 8844.13 meters, which was considered the Qomolangma's official height in China until 2005.

Version 3: This data had been adopted by the Chinese government till 2005.

Version 4: This number had been adopted as the height of Chomolungma in China and it had been used till 2005.

Version 5: This figure was considered the Qomolangma's official height in China until 2005.

In translating the words 数据、采用值 and 沿用, students who apply to China Translation Dictionary had much more choices than those who use traditional dictionaries as shown in the ensuing examples:

○学院开办的晚间短期课程约有500项，学制沿用全年三个学期的形式

More than 500 part-time evening subjects are planned for each of the three terms of the year.

○沿用旧制To follow an old course

○这个制度一直沿用到今天，即淘汰赛后，仅仅剩下十六支球队角逐世界杯。

This system is still used today, and only the sixteen teams left after elimination actually compete for the Cup.

○非营利组织会计也沿用与政府会计相似的会计模式。

Nonprofit accounting also follows a pattern of accounting that is similar to governmental accounting.

○在这些拥挤的社区中，人们仍然沿用自己故乡的语言、制度以及社会传统。

In these packed communities, the people clung to their native languages, institutions and folkways

○喜欢它的人则说它是要一直沿用的。

Its most ardent admirers would say that it will be around forever

○一八六八年在关于建立加利福尼亚大学的特许状中，就要求该校招收女生，这是沿用在此前二十年中中西部一些州办大学所采取的办法。

The 1868 charter of the University of California required it to admit women, following the lead of several midwestern state universities founded in the previous two decades

○1967年，瑞士人已经获取了数字技术的专利权，但瑞士的钟表制造商却对此不屑一顾，依然钟情于沿用了几十年的传统制造方法。

In 1967, when the digital technology was patented by the Swiss, they rejected it in favor of the traditional techniques they had been using to make watches for decades.

○他沿用祖父的名字，被命名为理查。

He was called Richard after his grandfather.

○这个故事中的日期均沿用农历。

All the dates given in this story follow the lunar calendar.

○本法施行前设立的商业银行，其组织形式、组织机构不完全符合《中华人民共和国公司法》规定的，可以继续沿用原有的规定，适用前款规定的日期由国务院规定。

If a commercial bank, which was set up before the promulgation of this law has its organization form and setup not entirely in accordance with the stipulations of the "Company Law of the PRC ", it may continue to operate under the original form and setup, and when the stipulations of the preceding clause should be strictly applied is up to the decision of the State Council.

Finally, the translation of the Proper Noun will be discussed. Short as it is, there are many Proper Nouns in the text, such as 珠穆朗玛峰、喜马拉雅山脉、干城章嘉峰、洛子峰、马卡鲁峰、中国科学院、中国国家测绘局、中国国家体委, etc. These terms should not be rendered in the method of transliteration or in accordance with Chinese Pinyin pronunciation. Instead, the students should resort to dictionaries or online resources to achieve acceptable versions, for example, Qomolangma (or Mount Everest), Himalaya, Kanchenjunga, Lhotse, Makalu, Chinese Academy of Sciences , the State Bureau of Surveying and Mapping, the State Physical Culture and Sports Commission. It should be stressed that 国家 is usually rendered as State in most Chinese-English dictionaries, which differs from the expressions given by some foreign media, say, the National Bureau of Surveying and Mapping of China, the National Physical Culture and Sports Commission of China.

From the above occurrences, it can be seen the corpus has more varieties than the dictionary. This is also reflected in the students' translated texts. Despite the grammatical errors and inadequacies made by corpus users, their translated texts have more varieties than dictionary users. It is also true in this case that some dictionary users render better translation than corpus users.

This, is believed, simply attests to the fact that translation is a very complicated activity which is closely related to students' linguistic competence, cultural awareness, etc. However, the general trend tells us that corpus users show more resourcefulness in their translated texts than most dictionary users.

5. Discussions

The experiment has given us some new insights into the translation teaching course and other related areas.

First, corpus proves more useful than bilingual dictionaries. In our experiment, it proves that corpus, to a certain extent, is a more useful tool than bilingual dictionaries. The corpus is based on a cohesive text while bilingual dictionary is based on word-for-word translation. With the help of a corpus, students can understand how a certain word is used in a text; at the same time, they can also understand better the subject matter. These advantages of corpus are something that bilingual dictionaries do not have. By using a corpus, translators can be more efficient in their queries of information. One can easily obtain large quantity of information in seconds by simply entering the keywords. Moreover, the large quantity of occurrences available in the corpus is also a strong point.

Secondly, students show more confidence when using corpus. It is also found that students show more freedom in their choice of words and expressions. In comparison to the limited translation equivalents provided by the bilingual dictionaries, corpus can provide a great number of different translations with only a click of the mouse.

Thirdly, the use of corpus shows little significance in helping students reduce grammatical errors. The number of grammatical errors made by corpus users is no less than those made by

dictionary users. This shows that the quality of translation is closely related to students' language ability. When students are incompetent in their language ability, translation is merely an impossible mission. Though all the participants in our experiment are third year students majoring in English, it is found that their target text production is not very satisfying in terms of the number of grammatical errors. In this specific experiment, it is found that the corpus does not prove to be very useful in this respect.

6. Conclusion

The current research aims at improving the current situation of translation teaching class through a corpus-based approach. Our research has resulted in some new findings, though not very significant. As to translation teaching classes, the implications are far reaching. Through this study it is indicated that the current translation teaching class needs reassessment of teachers' and students' roles in the classroom, redefining the teaching goal, readjustment of the translation teaching syllabus and other areas also need to be examined. Corpus here mainly opens for us a window to see more clearly all these problems. Despite its many weaknesses, it can be a very useful tool for both teachers and students to enhance the performance and quality of translation teaching classes.

Bibliography

- Baker, M. (1992) *In Other Words: A Course book on Translation*. London & New York: Routledge.
- Baker, M. (1993) *Corpus linguistics and translation studies: Implications and applications*. In M. Baker, C. Francis & E. Tognini-Bonelli (eds.), *Text and Technology: In Honor of John*

- Sinclair*. Amsterdam&Philadelphia: John Benjamins, 233-250.
- Baker, M. (1995) Corpora in translation studies: an overview and some suggestion for future research[J]. *Target* 7(2): 223-43.
- Baker, M. (1996) *Corpus-based Translation Studies. The Ahead[A]. in Harold Somers (eds.)*, Terminology, LSP Amsterdam: John Benjamins, 175—86.
- Bowker, L.&P. Bennison, (2003) Student Translation Archive. Design Development and Application. In Zanettin, F., Bernardini S. and Stewart D., 2003, (eds.) *Corpora in Translator Education*. Manchester: St Jerome,104-117.
- Kubler, N. (2003) Corpora and LSP Translation[A]. in Zanettin, F., Bernardini, S.& Stewart, D. (eds.), *Corpora in Translator Education*. Manchester: St.Jerome,
- Laviosa, S. (1998) Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* XLIII, 4. 557-570.
- Li, Defeng. (2003) Translator Training: What Translation Students Have to Say. *Meta: Translators' Journal*. 47:4, 513-531.
- Maia, B. (2003) Some languages are more equal than others. Training translators in terminology and information retrieval using comparable and parallel corpora. In Zanettin, F, Bernardini S. and Stewart D., 2003, (eds.) *Corpora in Translator Education* Manchester: St Jerome, 43-53.
- McEnery, A.&A. Wilson. (1996) *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Mu Lei. (1999) Translation Teaching in China. *Meta* XLIV 198-208.
- Munday, J. (1998) A computer-assisted approach to the analysis of translation shifts. *Meta* XLIII, 4. 542-556.
- Munday, J. (2001) *Introducing Translation Studies: theories and applications*. London: Routledge.

- Nord, C. (2001) *Translating as a Purposeful Activity: Functionalist Approaches Explained*.
Shanghai: Shanghai Foreign Education Press
- Toury, G. (1995) *Descriptive Translation Studies and beyond*. Amsterdam & Philadelphia:
Benjamin's.
- Varantola, K. (2003) Translators and Disposable Corpora in Zanettin, F., Bernardini S. and
Stewart D (eds.) *Corpora in Translator Education*. Manchester: St Jerome, 55-70.
- Wilkinson, M. (2005) Using a Specialized Corpus to Improve Translation
Quality. *Translation Journal* Vol. 9:3.
- Williams, J.&A. Chester man. (2002) *The Map: A Beginner's Guide to Doing Research in
Translation Studies*. Manchester, UK: St. Jerome Publishing.
- Wills, W. (2001) *The Science of Translation: Problems and Methods*. Shanghai: Shanghai
Foreign Education Press
- Xiao, Z.&McEnery, A. (2004) Parallel and comparable corpora: What are they up to? Available
online at: <http://www.lancs.ac.uk/postrg-ad/xiaozlpapers/corpora%20and%20translation>
- 陈瑞清. (2003) 语料库翻译学: 英汉翻译的研究与应用初探. 第八届口笔译教学研讨会论文集.
台湾翻译学学会7-1—7-28
- 丁信善, (1998) 语料库语言学的发展及研究现状. *当代语言学*(1): 4-12
- 冯跃进, 陈伟 (1999) 汉语“副职”英译的语料库调查研究. *外国语*(2): 43-49
- 柯飞. (2002) 双语库: 翻译研究新途径. *外语与外语教学*(9): 35-39
- 刘康龙, 穆雷 (2006) 语料库语言学与翻译研究. *中国翻译*(1): 59-64
- 穆雷. (2004) 翻译教学发展的途径. *中国翻译*(5): 25-26
- 秦洪武, 王克非. (2004) 基于语料库的翻译语言分析——以“so---that”的汉语对应结构为例. *现代
外语*(1): 42-48

- 隋桂岚, 刘励. (2004) 关于大型文学语言语料库的构建. *外语电化教学*(2): 37-39
- 王建新. (1998) 语料库语言学发展史上的几个重要阶段. *外语教学与研究*(4): 52-58
- 王克非. (2003) 英汉/汉英语句对应的语料库考察. *外语教学与研究*(6):410-416
- 王克非. (2004b) 双语平行语料库在翻译教学上的用途. *外语电化教学*(6):27-32
- 王克非. (2004a). *双语对应语料库研制与应用*. 北京:外语教学与研究出版社。
- 杨惠中. (2002) *语料库语言学导论*. 上海:上海外语教育出版社。
- 杨建玫. (2002) 警察与赞美诗的语料库检索分析. *四川外语学院学报*(3):56-59
- 于连江. (2004) 基于语料库的翻译教学研究. *外语电化教学*(2): 40-44

Appendix (Text used in the research)

珠穆朗玛 —— 雪山女神

珠穆朗玛峰位于中国和尼泊尔两国边界，海拔8844.43米，是喜马拉雅山脉的主峰，也是世界第一高峰，被誉为“万山之尊”、“地球之巅”、“世界第三极”。

珠穆为藏语“女神”之意，朗玛是“第三”之意，形容其柔美、静穆，宛如云中仙子。峰顶终年积雪，在劲风的吹拂下，飘飘洒洒，好像女神的面纱，给人以美丽、肃穆、神圣的感觉。

珠穆朗玛整个山体呈金字塔状，威武雄壮，昂首天外，四周地形极其险峻，环境异常复杂。在它周围20公里的范围内，群峰林立，山峦叠障。海拔7000米以上的高峰就有40多座，较著名的有干城章嘉峰（世界第三高峰）、洛子峰（世界第四高峰）、马卡鲁峰（世

界第五高峰)等,形成了群峰来朝,峰头汹涌的场面,壮观无比。

1975年,中国科学院、中国国家测绘局和中国国家体委组成联合登山科学考察队,登上珠峰之巅。当时中国科学家测定珠峰高度为8848.13米。这一数据一直作为中国对珠峰高度的采用值沿用至2005年。2005年10月9日,中国国家测绘局公布了珠穆朗玛峰新的高度为8844.43米。