

Parallel Corpus in Translation Studies: An Intercultural Approach

Xiaoping (Isadora) Jiang, Guangzhou University

Josta van Rij-Heyligers, University of Auckland

The international symposium on
Using Corpora in Contrastive and Translation Studies
Hangzhou China, 25th – 27th September 2008

Abstract:

This paper explores the potential use of parallel corpus in translation studies from an intercultural perspective. To achieve this objective, the paper first investigates the early development of parallel corpora. This is followed by an examination of current English and Chinese corpora and their applications in translation. To illustrate the practical use of parallel corpus in translation, the paper makes a comparative study of some source texts and their translations both in English and Chinese at lexical, syntactical, and discourse levels. By using these examples, the paper attempts to explore the possibility of providing an intercultural dimension in the translation classroom and address the value of intercultural knowledge in the translation process and foreign language studies. Lastly, the exploitation of parallel corpora for teaching translation as well as current limitations is discussed.

Key words: parallel corpus; translation studies; intercultural competence; English and Chinese corpora

Introduction

The corpus-based approach to studying translation has become popular over the last decade, with a wealth of data now emerging from studies using parallel corpora, multilingual corpora and comparable corpora. Corpora, particularly parallel corpora, prove to be valuable sources of information in translation research and education. As Varantola (1997) indicated translators can spend up to 50% of their time on consulting reference materials. In this context, the use of computer-based bilingual corpora can enhance the speed of translation as well as its quality, for they enable more native-like interpretations and strategies in source and target texts respectively (Aston 1999). They also help trainee translators to become aware of general patterns and preferred ways of expressing things provided the corpora of texts are varied and large enough (Zanettin 1998). As such, corpora function as helpful resources to developing an understanding of conventions and norms in different languages. In the design and use of translational corpora it is important, however, that the social and cultural contexts in which translations are produced and interpreted are not neglected in order to “provide a framework within which textual and linguistic features of translation can be evaluated” (Bernardini & Zanettin 2004: 60).

This paper explores the potential use of parallel corpus in translation studies from an intercultural perspective. To achieve this objective, the paper first investigates the early development of parallel corpora. This is followed by an examination of current English and Chinese corpora and their applications in translation. To illustrate the practical use of parallel corpus in translation, the paper presents a comparative analysis of some source texts and their translations both in English and Chinese at lexical, syntactical and discourse levels. By using these examples, the paper explores the possibility of providing an intercultural dimension in the translation classroom and addresses

the value of intercultural knowledge in the translation process and foreign language studies. Lastly, the exploitation of parallel corpora for teaching translation and the current limitations are discussed is discussed.

English and Chinese corpora and their recent applications in translation

Since the 1980s, corpus linguistics has gained in increasing popularity. What is corpus linguistics? Why is it becoming so prominent? How is it applied to translation? These questions beg some interpretations and discussion. According to Michael Stubbs (2001: 151), “Corpus linguistics [...] investigates relations between frequency and typicality, and instance and norm. It aims at a theory of the typical, on the grounds that this has to be the basis of interpreting what is attested but unusual” . To put it simply, corpus linguistics refers to the study of linguistic phenomena through large collections of machine-readable texts: corpora (any text in written or spoken form). It is based on empirical evidence in how language is actually used. Corpus linguistics is, therefore, of great importance in lexicography and dictionary design, grammar and, increasingly, English language teaching and second language acquisition.

A landmark in modern corpus linguistics was the publication of *Computational Analysis of Present-Day American English* by Henry Kucera and Nelson Francis in 1967, a work based on the analysis of the Brown Corpus, a carefully compiled selection of current American English. The book comprises up to about a million words drawn from a wide variety of sources. Kucera and Francis subjected it to a variety of computational analyses, from which they compiled a rich and variegated opus, combining elements of linguistics, language teaching, psychology, statistics, and sociology.

A further key publication was Randolph Quirk’s “Towards a description of English Usage” (1960,

Transactions of the Philological Society, 40-61) in which he introduced The Survey of English Usage. Shortly thereafter, Boston publisher Houghton-Mifflin approached Kucera to supply a million word, three-line citation base for its new American Heritage Dictionary (AHD), the first dictionary to be compiled using corpus linguistics. The AHD made the innovative step of combining prescriptive elements (how language should be used) with descriptive information (how it actually is used).

Other publishers followed suit. The British publisher Collins' COBUILD monolingual learner's dictionary, designed for users learning English as a foreign language, was compiled using the Bank of English.

The Brown Corpus has also generated many similarly structured corpora: the LOB (Lancaster-Oslo-Bergen) Corpus (1960s British English), Kolhapur (Indian English), Wellington (New Zealand English), Australian Corpus of English (Australian English), the Frown Corpus (early 1990s American English), and the FLOB (the Freiburg-LOB) Corpus (1990s British English), and an update of the LOB corpus in the early 1990s. Other corpora represent many languages, varieties and modes, and include the International Corpus of English (ICE), and the British National Corpus, a 100 million word collection of a range of spoken and written texts, created in the 1990s by a group of publishers, universities (Oxford and Lancaster) and the British Library. For contemporary American English, work has stalled on the American National Corpus, but the 360-million-word Corpus of Contemporary American English (COCA) (1990-present) is now available ([Wikipedia 2008](#)).

As investigated above, the construction and use of English language corpora dominates the research of corpus linguistics, but corpora of other languages like French, German, Chinese and Japanese have become available, and they are good additions to the corpus-based language studies

and literature.

In China, the history of the development of Chinese corpora can be roughly divided into three stages: from very early 20th century to 1980s; in the period between 1980s and the early 1990s; and since the middle of 1990s. The first stage is seen as the age of pre-computer in China. Chinese printed texts used to be manually collected as corpora, on which frequency of Chinese characters can be counted by hand. Corpora in this stage were used to learn about the actual usage of Chinese characters in a real world context according to statistical data. This work is of great value to help the compilation of Chinese textbooks used for Children's learning to read and write Chinese characters in primary schools (Zhan et al. n. d).

The second stage is marked by the use of a computer to store and process digital documents. In general, the size of a Chinese corpus in this period amounts to millions or even ten millions of Chinese characters. With a view to counting the frequency of Chinese words rather than characters in a real world context, researchers segmented Chinese sentences into word sequences by hand with the aid of computer. Contrary to corpora in the first stage, the basic unit of corpora in the second stage is word instead of Chinese character. The main applications on Chinese corpora in this stage include compiling Chinese word frequency dictionaries; selecting most frequently-used words for use in Chinese textbooks; and drafting the specification for Chinese word segmentation, which was revised and finally issued as the national standard (numbered GB-13715) in October of 1990. The national standard, titled as *The Segmentation Criterion for Modern Chinese Used for Information Processing*, is the first guideline for automatically segmenting Chinese written language (Zhan et al. n. d).

Since the middle of the 1990s, computer and software on natural language processing have been used more broadly in the development of Chinese corpora, including not only written text material

but also colloquial Chinese, with detailed annotations. Some new trends in this development can be outlined by the representative Chinese corpora: very large and detailed annotated corpus, multimedia corpus, cross-language corpus, and special purpose corpus (Zhan et al. n. d).

The above examination is concerned with the main developments of English and Chinese corpora, which occurred much earlier than the advance of corpus-based approaches to studying translation.

These approaches, however, have recently developed at a much accelerated pace. And with increasing data obtained from corpus research and the development of parallel corpora, multilingual corpora and comparable corpora, corpora are becoming useful resources in the translator training classroom and translation studies.

The corpus-based approach to translation studies emerged as a new translation research paradigm in the 1990s. Its significant contribution is the research on the universals of translation, which indicate the typical features of the translated text rather than the original text. These translation universals are represented by simplification, explication, standardization, regularization and centralization (Hu 2004).

Parallel corpora, meaning large collections of texts in two languages, have been a key focus of non-English corpus linguistics, mainly because this type of corpora is an important resource for translation and contrastive studies. As Aijmer and Altenberg (1996: 12) observe, parallel corpora “offer specific uses and possibilities” for contrastive and translation studies: they give new insights into the languages compared - insights that are not likely to be noticed in studies of monolingual corpora;

- they can be used for a range of comparative purposes and increase our knowledge of language-specific, typological and cultural differences, as well as of universal features;
- they illuminate differences between source texts and translations, and between native and

non-native texts;

- they can be used for a number of practical applications, e.g. in lexicography, language teaching and translation.

In this paper, the potential value of parallel corpora in translation studies and concerns are explored. Before doing so, however, it is necessary to provide case studies of some source texts and their translations both in English and Chinese to highlight significant cultural differences embedded in the two languages with a main view to stimulating intercultural awareness and developing a high degree of intercultural sensitivity.

Case studies of corpus: Some source texts and their translations both in English and Chinese

In this section, Chinese and English are chosen as the two languages for case studies from an intercultural perspective. The contrastive studies are to be made at lexical, syntactical, and discourse levels. It is thus of great use to point out some significant differences between Chinese and English in order to help translators do effective translation.

Pascale Fung (n.d.) identifies a few typical features of the Chinese language that makes it distinct from English. First, Chinese texts have no word delimiters. Second, Chinese part-of-speech classes are very ambiguous; many words can be adjective or noun, noun or verb; many adjectives can also act as adverbs with no morphological change. Third, Chinese words have little or no morphological information. There are no inflections for nouns, adjectives or verbs to indicate gender, number, case, tense or person (Xi 1985). Fourth, there is no capitalization to indicate the beginning of a sentence. Fifth, very few function words exist in Chinese compared to other languages, especially to English. Moreover, function words in Chinese are frequently omitted. Sixth, a large number of acronyms are used in Chinese, which means many single words in

Chinese can be translated into compound words in English. Seventh, the usage of idioms in Chinese is considerably more frequent than in English. These features of the Chinese language demonstrate that the Chinese texts in a corpus have fewer unique words than those in an English corpus.

Lexical level

Words are the basic units of meaning. Understanding the meanings of words is, therefore, critical to the sharing of meanings conveyed in verbal communication, especially when two different languages like Chinese and English are involved. Lexical meaning can largely be classified into two types: denotation and connotation. Denotation is the conceptual meaning of the word that designates or describes things, events or processes, etc. whilst connotation indicates the emotional or stylistic association that a word or phrase evokes in one's mind (Song 2004).

The meaning of words is personal and cultural. Each individual has his/her own sphere of experience. And his/her connotations of certain words may be based on his/her unique experience with the referent. The word 'snow' may evoke very different emotional and perceptual meanings to Eskimos who live in a snow world throughout the year and to people in Hainan Island, China, who live a tropical life.

Meanings are also culture specific. Since people of the same culture share much in physical environment, customs, traditions and other cultural heritage, they tend to have the same or similar understandings of what a word means and associative meanings derived from it, as represented by the word 'snow'. Both denotations and connotations are used in interpersonal communication. The understanding of meanings depends on the extent or degree to which the listener and speaker share the repertoire of meanings attributed to the employed codes.

Communicative problems often arise from connotative meanings. People tend to behave as if their connotative reactions to a word or concept are the same as reactions somebody else would have. Consider the word ‘dog’ in English and 狗 in Chinese. They can be said to have the same denotative meaning; however, people from different cultures may have different connotative reactions to the above two words. Dogs in the English language have very positive connotations such as being faithful and caring. Hence, there are many English phrases as shown in the following table, which are associated with luck and happiness. By contrast, dogs in the Chinese language have negative connotations; thus, Chinese use the word ‘dog’ in a derogative way, generating many phrases to show their dislike of dogs, as shown below.

Table 1: Different connotations of ‘dog’ in Chinese and English and their translations in the two languages.

Dog

English	Chinese Translation	Chinese	English Translation
A homeless dog	丧家之犬	狗腿子	lackey
Love me, love my dog.	爱屋及乌	狗养的, 狗崽子	bastard
Top dog	最重要的人物	狗胆包天	Monstrous audacity
Lucky dog	幸运儿	狗屁	Horseshit; rubbish
Dog-tired	精疲力竭	狗嘴里吐不出象牙	A filthy mouth can't utter decent language.
To die like a dog	可怜地死去	狗头军师	A person who

To help a lame dog over a stile	助人于危难	狗仗人势	offers bad advice A dog bites on the strength of his master's backing.
------------------------------------	-------	------	---

An examination of the above table reflects the consideration of cultural differences given in the translation of the word 'dog' in the two languages. Without taking into account the different cultural associations of the word 'dog' in translation, misunderstanding would arise in intercultural communication.

Syntactical level

Learning a language is much more than learning the vocabulary. The good command of English vocabulary may not guarantee the learners to combine the English words to form meaningful sentences. Many university students in China have memorized over 5000 words but they cannot write idiomatic sentences, let alone paragraphs. Linguistic and cultural barriers are the main barriers for Chinese learners since Chinese and English belong to two different language families and have different language and cultural systems. Additionally, the influence of mother tongue remains strong in the foreign language studies. Therefore, it is necessary to investigate the differences between sentence structures in Chinese and English.

Chinese is a typical language with a paratactic sentence structure, meaning the construction of sentences or clauses through semantic ties rather than connective words; whilst English is one with a hypotactic sentence structure, indicating the construction of sentences or clauses through linguistic forms including conjunctions, adverbs, and prepositional phrases. The Chinese

sentences are organized according to the meaning. If meaning is understandable, the sentence is acceptable. On the contrary, the English sentences are organized according to the form or grammatical framework, with the formal relations being stressed. Without the proper forms, meanings cannot be clearly expressed in English (Song 2004).

Being dependent on linguistic forms, English is rigid in the construction of sentences whereas Chinese is much more flexible. This is particularly true in ancient Chinese poems. The following is the poem written by a famous ancient Chinese poet Ma Zhiyuan (马致远) and its English translation (Jia 1997).

古藤, 老树, 昏鸦,
 小桥, 流水, 人家.
 古道, 西风, 瘦马,
 夕阳西下, 断肠人在天涯。

*Crows hovering over rugged old trees wreathed
 with, rotten vine – the day is about done.
 Yonder is a tiny bridge over a sparkling stream,
 and on the far bank, a pretty little village.
But the traveler has to go on down this ancient road,
the West wind moaning, his bony horse groaning,
 Trudging towards the sinking sun, farther and farther
 away from home.*

A careful comparison of this Chinese poem and its English translation clearly reveals the paratactic structure in Chinese and the hypotactic structure in English. In the Chinese poem, the first three lines comprise only noun phrases. No subjects are used in the last two lines. The Chinese language is such a system that it offers the poet enormous freedom of creation. Similarly, the readers can enjoy the pleasure of joining the poet in his or her imagination. The meaning is as much embedded in the language as in the readers' recreation. The setting of the poem is so vividly described with the noun phrases that any addition of connectives would be superfluous and destructive. However, in the English version the rules of writing are different. Connectives, prepositions, verbs, pronouns, etc. have to be used to indicate the relationships between each element in the poem so that the meaning will be clear to the reader. Linguistic forms ensure the writer construct the sentences in a logical way that meaning is most clearly conveyed to the reader. The syntactic differences between the two languages also uncover different cultural features. Chinese tends to be reserved and indirect whilst English-speaking people tend to be more open and direct in communication. Therefore, in translation, not only linguistic forms but also cultural differences should be taken into serious consideration.

Discourse level

Discourse is a piece of speech or writing, which may be as short as an exchange of greetings, a paragraph, or as long as a lecture or an essay (Song 2004). People from different cultural backgrounds are significantly different in the way they organize and deliver their spoken or written discourse, which tends to cause misunderstanding in communication. English and Chinese speakers, under the strong influence of their native languages, are likely to express the same idea in different ways. This is especially true in written discourse. Discourse is the highest level at which

we may find cultural deposits, which usually take the form of the discourse pattern and the stylistic features of a specific genre. Hoey (1983) identifies three discourse patterns in English: the Problem-solution Pattern, the General-particular Pattern and Matching Pattern. Chinese students tend to be weak at the General-particular Pattern and often confuse English readers with an Implication Pattern in their composition. This means that source patterns make the target text strongly assertive, so that the target language readers have to adapt their thinking to the original writer.

In Chinese-English translation, the translator chooses to change the discourse pattern of the source language text in order to make the translation correspond to the target language discourse norms, particularly in the translation of advertisements. The changes are often made in the structure and genre of the text. Traditional Chinese texts are characterized by a four-stage pattern: introduction, development, transition and conclusion. Although this pattern is not seen as the only one guidance in today's Chinese writing, its influence is still dominant (Li 2003).

The following example, extracted from a review on the Chinese classic, *The Dream of the Red Mansions*, illustrates the typical Chinese writing pattern including introduction, development, transition and conclusion. The translator deletes the introduction and the conclusion (the underlined parts in Chinese and brackets in the English version) and keeps only the information that fits into the overall text of the review.

小说的艺术表现可以说是达到了出神入化的境界。严谨、缜密的结构，生动准确的个性化语言，特别是鲜明的人物形象，都是非常杰出的。书中出现的人物，粗计多达四百多人。不仅主角贾宝玉、林黛玉和其他十多名主要人物成为人们熟知的艺术典型，而且许许多多次要人物，有的甚至是一笔带过的，也都是形象鲜明，栩栩如生。小说的艺术表现，达到

如此高超的水准，在世界文学名著中，也是极为罕见的。

[It can be said that the artistic presentation of the novel has reached the acme of perfection.] The novel is outstanding for its brilliantly balanced structure, a lyrical yet precise prose style and rich characterization. Although there are more than four hundred characters, the principal characters number fewer than twenty with Jia Baoyu and Lin Daiyu pre-eminent. Among the array of secondary characters, even those appearing only briefly are clearly drawn and realistic. [The novel's artistic presentation has reached a high standard that is rarely seen among literary works worldwide.]

Another good example includes a short paragraph extracted from an English introduction to the Chinese university, Tianjin Normal University. This example illustrates

天津师范大学是一所综合性的重点高等师范院校，诞生于1958年，随着共和国的成长，她也历经磨砺，做过了三十多年的风风雨雨。三十多年来，一批批德才兼备的教学、科研和管理人员会集在这里，怀着振兴天津，振兴教育的希冀，在这块土地上默默地耕耘。春华秋实，硕果累累，两万多名教育教学人才走出学校大门，足迹遍及全国，桃李满天下。他们献身教育，殚思竭虑，为撑起教育的脊梁，托起明天的太阳。

In the above description of the university, there are many formal, descriptive words and long complex sentences. The language is too redundant and the style flowery. But simple language and plain style is what the native English readers and writers take for good writing in English except for literature creation. The main purpose of this introduction is to provide the audience with reliable information. The English translator, sticking to the 'KISS principle' in the English writing: keep it short and simple (Katan 2006), reduces the literary pomposity to simple exposition. The underlined parts of the Chinese text are either ignored or given a free translation to make it more to the point. To compare the following English version with the original Chinese version reveals the

different writing styles in the two languages.

Founded in 1958, the University has entered its fourth decade with a remarkable record of both hardship and achievements. It is now ranked as one of the key institutions of teacher training in China. [Along with the growth of the Republic, she has stormed 30 difficult years of wind and rain.] In the past thirty years, hundreds of talented teachers, researchers and administrators have gathered and worked here [in the hope of rejuvenating Tianjin and education, and they silently plowed on this piece of land] in a continuous endeavour to meet the ever-increasing demands for educators both in Tianjin and the rest of the country. [Flowers bloom gloriously in spring, and fruits hang heavy in autumn.] More than 20 thousand students have graduated from the University and are now teaching nationwide. [Their footprints are all over the country and they bear fruits like peaches and plums all over the world.] They have dedicated their wisdom and energy to the educational needs of the country [to prop up the backbone of education and the rising sun of tomorrow] in the firm conviction that the future of China lies in the education of the younger generation.

In brief, a careful analysis of the differences between Chinese and English at the lexical, syntactic and discourse levels from an intercultural perspective through these case studies shows that an effective translation can hardly be achieved without taking cultural factors into consideration. If parallel corpus could be fully applied in translation from an intercultural approach, translation could reach a new platform.

The next section is devoted to exploring how intercultural awareness could be realized in the translation classroom teaching.

Realizing intercultural awareness in the translation classroom

Translation is a form of intercultural communication. It is not merely a linguistic activity, but essentially a cultural communication. Translation involves at least two languages, which means dealing with two cultures since language is seen as a carrier of culture. Thus, the translator is not only a bilingual mediating agent but also a cultural mediator who facilitates communication, understanding, and action between people or groups from different language and cultural backgrounds.

Therefore, in the translation classroom, the teacher needs to remind his/her students of the important role of cultural mediator as translator. To become a skilled mediator, one must possess the following competences in both cultures (Taft 1981:73).

- Knowledge about society: history, folklore, traditions, customs; values, prohibitions; the natural environment and its importance; neighbouring people, important people in the society, etc.
- Communication skills: written, spoken, non-verbal.
- Technical skills: those required by the mediator's status, e.g. computer literacy, appropriate dress, etc.
- Social skills: knowledge of rules that govern social relations in society and emotional competence, e.g.: the appropriate level of self-control.

In order to play the role of mediator, "an individual has to be flexible in switching his cultural orientation" (Taft 1981:53). Therefore, a cultural mediator will have developed a high degree of intercultural sensitivity, awareness and competencies. As a cultural mediator, he or she will need to be a specialist in negotiating understanding between cultures. However, in the translation process, a translator tends to be influenced by his or her own beliefs, knowledge, attitudes and so

on. Thus, any translation will, to some extent, mirror the translator's own mental and cultural outlook, despite the best of impartial intentions (Katan 2006). Translators as cultural mediators should therefore be highly aware of their own cultural identity; and for this reason will need to understand how their own culture influences their perception of the original text.

The new global context calls for a new teaching approach to teaching translations. This change in the global context demands a privileging of the cultural function within traditional teaching of translation, that is, the realization of intercultural education in teaching translation. Translation teaching can no longer be regarded as a mainly linguistic task. Teachers are now required to implement intercultural education in translation teaching. They cannot be confined to the three traditional teaching models of translation in China: the translation skills-oriented model, the translation theories-oriented model; and the translation theory and practice model, for the main limitations of the three models lie in the treatment of translation as a cultural static phenomenon, independent of cultural interaction. These limitations prevent translation from serving as a bridge to effective communication of cultural values, especially in the context where there is the need for mediation between very diverse cultures. Hence, translation teaching should be treated as a cultural communication activity rather than purely a linguistic interaction (Jiang 2008) and that it should focus on developing students' intercultural competence and global awareness and promoting understanding and respect for different cultures. The question is whether corpus-based approaches using parallel texts can contribute to this development.

Potentiality and issues of the use of parallel corpus resources in translation

As Aduari Brezolin (2008) stated, it seems that contrary to practice many theorists in applied translation studies believe that translation work is unidirectional - translation into L1, as it is

assumed that most translators are not expert enough to translate L1 text adequately into L2.

However, such a view excludes trainee translators from developing language and cultural competence in both languages as in translating texts the focus is on making the target text culturally, lexically, syntactically and discursively comprehensible to their (L1) audience. To avoid such one-way linearity, corpora, in particular parallel corpora, can help the translator in training. As previously mentioned, parallel corpora are of practical use in language and translation classes. They help provide insight into the authentic use of a word or a collocation and discover their suitable or natural equivalent(s) in translation, information not always obtained from dictionaries. They may also raise awareness of cultural elements in the source and target text and hence may assist in developing students' intercultural skills and sensitivity.

To illustrate, the English-Chinese Parallel Corpus and on-line Concord Program

(<http://ec-concord.ied.edu.hk/paraconc/index.htm>) developed by Wang, Lixun of the Hong Kong Institute of Education offers trainee translators the opportunity to freely access on the web corpus files from English (nearly a million words) and Chinese origin (nearly half a million words) that have corresponding translated versions in Chinese and English. Students can perform concordances in English and the different versions of Chinese (traditional, simplified and pinyin).

Although some versions contain fewer corpus files, a search for the translation of a word will generate several whole paragraphs in which the word occurs in the source and target languages. This allows students to see the word in a context and observe similarities and any deviations in translations. It further allows them to assess the word at the lexical and syntactical levels, and to some extent at the level of discourse.

Yet, the use of parallel corpora is still in its infancy and presents some limitations for contrastive and/or translation purposes.

One of the limitations of parallel corpora is that compilers often choose texts based on technical criteria such as accessibility, availability and copyright issues. For example, the above mentioned corpus contains parallel texts that are largely sampled from the classics (or dated documents) since these source texts and translations are relatively easy to obtain and bear no copyright. But as previously suggested, languages and cultures are far from static - they change over time be it at a different pace. Hence, a corpus of classics may well provide valuable snapshots of language use (in the past) but may be of less relevance for translating modern literature, texts or documents as words, collocations and their context undergo changes or expansion of their denotative and connotative meaning. For example, the word '*cool*' as in *cool bag* or *cool weekend* in contemporary English often means 'good, exciting, trendy, fun, first rate', meanings not used in the past. The equivalent Chinese translation '*fei-chang hao, heng bang*' is certainly less appealing than '*ku*' to reflect the background era. A corpus of classical texts thus often fails to provide cultural knowledge of the language communities as they interact today. The society, relationships, attitudes and (linguistic) behavior in the era of Jane Eyre bear little resemblance with those of Britain today (although deep traces can still be found).

In addition, the criteria of 'convenience' often lead to corpus texts being sourced from governmental proceedings or legal and official statutes (such as those from the European Union (EU) or bilingual states like Canada) as they are relatively easy to access. Most parallel corpora compiled from such texts are between (Indo) European languages, such as English-Spanish and English-German. Few bilingual Asian-European corpora exist; even English-Chinese corpora are sparse in comparison. And when legal and official texts are incorporated, like in Wang's parallel corpus, the size of the corpus is often small (approximately 32000 words), which puts restrictions on the search for terms and collocations, and clouds their general pattern.

Another limitation is that when a bilingual corpus is compiled, a global language like English often provides more of the original texts (as is the case in Wang's corpus) than languages of less international status. This factor should be borne in mind when working with translation corpora as issues of equivalence are raised. For example, Zanettin (1998) observed that translated texts:

- do not represent the full range of linguistic possibilities of the target language;
- reflect the stylistic idiosyncrasies (such as errors and/or linguistic choices) of individual translators.

Malmkjær (1998) highlights a similar problem: a parallel corpus usually holds only one translation for each source text. In so doing, it neglects a crucial feature of translation work – being aware of and negotiating the differences that exist between multiple translations of the same text. She further points out the concern that the concordance lines generated for analysis often fail to provide sufficient linguistic context to examine whole-text features and semantic phenomena with the result that only partial aspects of translation behaviors are revealed, whilst others are obscured (especially at the discourse level).

Recent developments in parallel corpora are encouraging though. The growth of parallel corpora built and made available by higher education institutions in China, Hong Kong and other places is a sign of more Asian-European corpora to materialize in the near future.

Another positive development is the construction of corpora containing large numbers of texts and their multiple translations. Malmkjær (1998) proposed, as solution to the problems she observed, a combined methodology of using large, norm-oriented parallel corpora with smaller corpora composed of carefully chosen source texts and many translations, which would allow for exhaustive examinations of entire texts. This approach offers richer findings and better accuracy in translation as it enables the detection of both idiosyncratic and norm-ruled behavior as well as

allows for the triangulation and contextualization of findings. The construction of corpora containing large numbers of texts and multiple translations can be seen as an extension of Malmkjær's proposal, as such corpora not only help illustrate how translators have achieved equivalence under certain conditions, but also provide ample examples of translation strategies and, in the process, raise awareness of general patterns, which trainee translators can observe from the recurring linguistic choices made by the translators of these texts (Yang and Li 2003). The structure of the English-Norwegian Parallel Corpus (ENPC) (see Johansson 1998) further offers a good design for limiting the dominance of English texts as original. This corpus is constructed as follows:

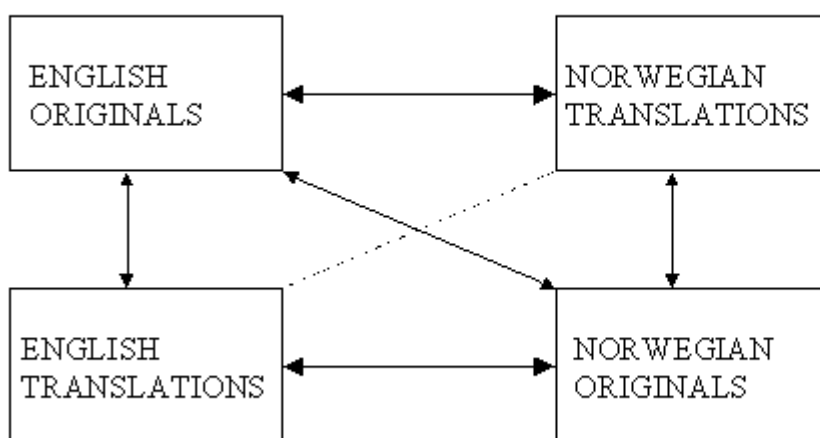


Figure 1. The structure of the ENPC (Johansson 1998)

The ENPC, a 2.6 million word corpus, consists of 100 original and 100 translated texts, equally sourced from fiction and non-fiction (even though this is less balanced) texts in the English and Norwegian language. The size of the boxes in Figure 1 represents the number of texts being nearly equal. Although this design poses a few problems to compilers as more English texts have been

translated in Norwegian than vice versa, it also has its strengths. As Johansson mentions, it enables (trainee) translators to perform contrastive studies using the parallel original texts (strong diagonal line) or the original texts and their translations (solid horizontal lines), or to perform various types of translation studies focusing on translation problems considered from either language (again the solid horizontal lines), deviations of translated texts as compared with original texts in the same language (strong vertical lines), and general features of translated texts (broken diagonal line).

It is important to have translations of source text and back translations of translated text, not only for accuracy but also for increasing understanding of similarities and differences at the cultural level. Such corpora could be helpful tools for developing intercultural skills in translation and communication. The design of the English-Norwegian corpus could function as a model for making this happen.

From a technical view, investigations that help with the mining of large texts and their translated versions from the Web (an almost unlimited source for texts and increasingly parallel texts) deserve special attention as they make the creation of large corpora for translation a reality. For example, Ying Zhang, Ke Wu, Jianfeng Gao, and Phil Vine (2006) noted that access to some parallel corpora is by subscription or license fee only. Others are domain specific, such as the European parliament proceedings parallel corpus (1996-2003) in eleven EU languages, and/or have been manually constructed. This limits their scope and practical use and any repeat of their method would be time consuming and costly. These authors instead used a method (a *k*-nearest-neighbors classifier) of multiple features that enabled them to identify, and automatically collect, parallel texts from the Web. Their evaluation of the obtained data set of 6500 Chinese-English candidate parallel pairs obtained a precision rate of 95% and a recall rate of 97%, making the method robust and promising for the future development of large parallel corpora from

different language systems.

Despite these developments, for most teachers of translation the issues surrounding the construction and use of parallel corpora are still far removed from their classroom practice. Nevertheless, bilingual and multilingual corpora can be successfully used in classroom activities and the literature provides some useful examples of how it can be implemented (see for example Zanettin 1998) so as to enhance trainee translators' linguistic and cultural competence as well as their translation skills such as cross-linguistic mediation and accuracy in text production. The challenge is for corpus researchers and designers to get teachers motivated by demonstrating their effectiveness. Important criteria here are likely to be simple use, easy access, high relevance and quality of results. This is not always possible especially since copyright issues usually limit the use of corpora to research activities (e.g. ENPC).

Conclusion

Bilingual (or multilingual) parallel corpora provide a new approach to translation studies. Clearly, they can be very useful resources and a practical tool in translation, for a parallel corpus can be used to do contrastive linguistic research, to analyze translational transformation or translators' style, to enrich a bilingual dictionary, and to help in translation training.

The main functions of a bilingual parallel corpus can be summarized as follows: it can provide various bilingual translation examples through concordance, many bilingual translation examples for oft-used structures so as to explain and mimic practice, and multiple translation versions for one original to compare. These functions cannot be fulfilled by other textbooks and dictionaries. Chinese-English parallel corpora, for example, represent potential key resources for

Chinese-English cross-language information processing, Chinese-English bilingual lexicography, and Chinese-English language research and teaching. They could be further exploited in translation between Chinese and English.

Having parallel corpora available, however, only forms a basis for the research of linguistics and translation. In this respect, much room for future work remains to be done. In the field of corpus-based translation studies, the properties of translated texts have been studied and compared to original text production within a language (Baker 2004). Such studies can yield interesting insights not only into the linguistic differences between translated and original texts, but also into the intercultural elements underlying translation. Then again, how to apply a bilingual parallel corpus such as Chinese-English in the translation teaching classroom, especially from an intercultural perspective, remains an issue to be addressed.

As discussed earlier, translation is a form of intercultural communication. Cultural factors needs to be taken into account in both translation and the teaching of translation. We could certainly use a bilingual parallel corpus to perform contrastive studies of two languages and cultures in translation studies. By analyzing different translation versions of the same original text, intercultural education could be incorporated in the translation teaching classroom. Corpus-based translation studies are still in its infancy since the field appeared only in the early 1990s. This leaves much space for further explorations. Undoubtedly, parallel linguistics has bright prospects for further application. It also has potential value for broadening our horizons of translation studies particularly in China.

References

Aijmer, K., Altenberg, B., and Johansson, M. (eds.) (1996). *Languages in contrast: Papers from a*

Symposium on text-based cross-linguistic studies, Lund: Lund University Press.

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

Baker, Mona (2004). 'A corpus-based view of similarity and difference in translation'.

International Journal of Corpus Linguistics 9 (2), 167-193.

Bernardini, S. & Zanettin, F. (2004). 'When is a universal not a universal? Some limits of current corpus-based methodologies for the investigation of translation universals'. In A. Mauranen, & P.

Kujamaki (Eds.), *Translation universals: Do they exist?* (pp. 51-62). Amsterdam: John

Benjamins.

Brezolin, Aduari (2008). 'Disposable corpus in translator training: Translating medical abstracts into L2'. *Paper presented at the 8th Teaching and Language Corpora (TaLC 8) Conference*, July 2008, Lisbon, Portugal.

Hall, Edward T. (1982). *The hidden dimension*. New York: Doubleday.

Hoey, M. (1983). *On the surface of discourse*. London: George Allen & Unwin (Publishers) Ltd.

Hu, Xiaoyao (胡显耀). (2004). 语料库翻译研究与翻译普遍性, <<上海科技翻译 >>, 第4期.

Jia, Yuxin. (1997). *Intercultural communication*. Shanghai: Shanghai Educational Publishing House. (in Chinese)

Jiang, Xiaoping (2008). 'The cultural mediation function of translation teaching'. *Foreign Languages and Their Teaching*, 7. (in Chinese)

Johansson, Stig (1998). 'On the role of corpora in cross-linguistic research'. In S. Johansson and S. Oksefjell (Eds.) *Corpora and cross-linguistic research: Theory, method and case studies* (pp. 3-24). Amsterdam: Rodopi.

Katan, D. (2006). *Translating cultures: An introduction for translators, interpreters and mediators*. Shanghai: Shanghai Foreign Language Education Press.

- Li, Yunxing. (2003). 'Cultural factors in inter-cultural translation'. In Guo Jianzhong's (Ed.) *Culture and translation* (pp. 184-196). Beijing: China Translation & Publishing Corporation.
- Malmkjær, Kirsten (1998). 'Love thy neighbour: Will parallel corpora endear linguists to translators?' *Meta*, 43 (4), 534-541.
- Pascale Fung. (n.d.). *Compiling bilingual lexicon entries from a non-parallel English-Chinese corpus*. Retrieved 18 July 2008, from <http://www.aclweb.org/anthology-new/W/W95/W95-0114.pdf>
- Song, Li. (2004). *Gateway to intercultural communication*. Harbin: Press of Harbin Institute of Technology.
- Stubbs, Michael (2001). 'Texts, corpora, and problems of interpretation: A response to Widdowson'. *Applied Linguistics* 22 (2), 149-172.
- Taft, R. (1981). 'The role and personality of the mediator', in S. Bochner (Ed.), *The mediating person: bridges between cultures*, pp. 53-88. Cambridge: Schenkman.
- Varantola, K. (1997). *Translators, dictionaries and text corpora*. Retrieved 16 March 2008, from <http://www.sslmit.unibo.it/cultpaps/varanto.htm>
- Wikipedia (2008). *Corpus linguistics*. Retrieved 19 July 2008, from http://en.wikipedia.org/wiki/Corpus_linguistics
- Xi, Zhu De (1985). *Yufa da weng - discussions on linguistics*. Hanyu Zhi Shi Cong Shu. Beijing, China: Shang Wu Yin Shu Guan. (in Chinese)
- Zanettin, Federico (1998). 'Bilingual comparable corpora and the training of translators'. *Meta*, 43 (4), 616-630. Retrieved 16 March, 2008, from <http://www.erudit.org/revue/meta/1998/v43/n4/004638ar.pdf>
- Zhan Weidong, Chang Baobao, Duan Huiming, and Zhang Huarui (n.d). *Recent developments in*

Chinese corpus research. Retrieved July 20, 2008, from http://www.tokuteicorpus.jp/result/pdf/2006_003.pdf

Zhang Ying, Wu Ke, Gao Jianfeng, and Vine Phil (2006). 'Automatic acquisition of Chinese–English parallel corpus from the Web', in M. Lalmas et al. (Eds.): *Advances in information retrieval*. Proceedings of the 28th European Conference on IR Research, ECIR 2006, Lecture Notes in Computer Science 3936, London, UK, April 10-12, 2006, (pp. 420–431). Heidelberg: Springer-Verlag.