

Specialized Comparable Corpora and Pragmatic Translation

Xia Yun Li Defeng

Qufu Normal University Shandong University

Abstract: This paper reports on the construction and use of specialized corpora in a comparative study of advertisements translated from Chinese into English and their comparable English texts, aiming to show whether and how specialized comparable corpora can be used to inform pragmatic translation. Ten translated advertisements of Chinese law firms and ten original ones from English-speaking countries are collected to compile the corpora, and Wordsmith tools is employed in the study. The result shows that differences between the translated texts and their comparable texts in the informative and vocative functions are manifested in aspects like informativity, point of view, and general style. On basis of this study, the author suggests that comparative studies with the help of comparable corpora prove a useful means in translation evaluation.

Keywords: pragmatic translation; comparable corpora; informative function; vocative function

1. Introduction

Pragmatic translation is essentially a decision-making process in which to achieve functional equivalence is the first priority, and where necessary, there is even need for stylistic recreation on the translator. Since pragmatic texts generally belong to what Newmark defines as ‘informative’ and ‘vocative’ texts (Jia, 2004:3), in pragmatic translation, the purpose of the overall translational action is to produce a target text that performs equivalent informative and vocative functions as the source text so as to meet the target audience’s expectations. For this purpose, the translator has to be familiar with the genre conventions that the target text is to conform to. If the text conforms to

the conventional patterns, the text form will not attract the readers' attention, which allows for an easier processing of the information contained in the text. On the other hand, if a text shows strange, unconventional form patterns, the audience may wonder why the author chose these original forms and whether they are meant to convey an extra amount of information (Nord, 2007). This paper reports on the construction and use of specialized corpora in a comparative study of English advertisements translated from Chinese and their comparable original English texts, aiming to show whether and how specialized comparable corpora can be used to inform pragmatic translation.

2. Comparable texts and Comparable corpora

The comparable corpus we talk about here refers to monolingual comparable corpus which 'consists of a corpus of translations and comparable non-translations in the same language' (Olohan, 2004:35). The authentic, non-translated texts are chosen from the target-text repertoire and represent the genre the target text is supposed to belong to. They constitute an important type of 'auxiliary texts' (see Nord, 2007:20) which serves as a source of cultural and linguistic information for translators. The first corpus of this kind, designed and constructed under the direction of Mona Baker, was the Translational English Corpus (TEC). It consists of a corpus of translations in English (from a range of different source languages) and is usually used with a comparable corpus of non-translations, which is a subcorpus of the BNC.

The aim of most research using a comparable corpus of this kind is in capturing 'patterns which are either restricted to translated text or which occur with a significantly higher or lower frequency in translated text' (Baker 1995:235), which can help find out about 'the nature of translated text in general and the nature of the process of translation itself' (ibid.: 236). Baker

(1996) posits a number of features (e.g. explicitation, simplification, normalization) as ‘universal features of translation’. Comparative study of translated and non-translated texts has proved useful in investigating the nature of translation process, translators’ styles, and translation norms in specific socio-cultural contexts. Pragmatic translation, however, can also benefit from it in that translators may learn to analyze the culture-specific features of textual and other communicative conventions in two cultures. This ability might be called ‘contrastive text competence’ (see Nord, 2007:19). Furthermore, through the quantitative analysis of particular lexical and grammatical features in the translated and non-translated texts, it is possible to investigate what characterize the translated texts, or the ‘translationese’.

The corpus-based approach makes the task of translation evaluation somewhat less difficult by removing a great deal of the subjectivity, and by providing the translator with improved access to the appropriate conceptual and linguistic information of a specialized subject field as documented by experts in that field (Bowker, 2001). It is the aim of this paper to demonstrate that specialized comparable corpora have a role to play in the assessment as well as improvement of translations.

3. Selecting texts

For the present study, collections of ten translated English advertisements of China’s law firms and ten non-translated ones from English-speaking countries are employed to construct our mini-corpus, with 7 translated ones from the U.S. and 3 others from U.K. All texts are collected through World Wide Web and saved in TXT format required by the corpus-processing software Wordsmith Tools. The two collections are rather small, with only 3,143 tokens in the translated texts and 2,584 in the non-translated. They are comparable in terms of corpus size and text

category (advertisements), and both are intended to perform the same informative and vocative functions for receivers familiar with the English language and culture.

The data are lemmatized, enabling the treatment of inflected forms of a word as belonging to the same base form or lemma, and this is required for a more accurate word frequency analysis as well as keyword analysis. My purpose is to find out how the two collections differ in their textual functions, using a few basic operations of Wordsmith computer program. However, it is clear that with such small collections as the ones used here, any conclusions that are drawn must be deemed to be tentative, and the discussion which follows is mainly illustrative, which is designed to show the potential usefulness of a comparable corpus in pragmatic translation evaluation.

4. Data analysis

4.1 General statistics and text style

The style of a text is one manifestation of its function in that different styles might produce different effects in the addressee. For example, formal style often implies seriousness and authority, while informal style is preferred on casual occasions and implies intimate relationship between the two sides. Wordsmith Tools provides the statistical information of the two collections of texts, which can give us a general idea of the style of the texts being investigated, as illustrated in Table 1. The Type/Token Ratio of translated texts is 29.62, much lower than non-translated ones, indicating a higher degree of lexical repetition in the former and greater lexical variation in the latter. The number of sentences in both collections is almost the same, but the average sentence length in translated texts is much higher than that in the non-translated ones. Generally, use of long and complex sentences is an indication of more formal, serious and less emotional style, while shorter and simple sentences often give an impression of a concise, casual and dialogic style. This

dialogic feature in the non-translated texts implies the addressers' intention to shorten the psychological distance between the potential addressees and themselves, which contributes to the 'appellative' or 'vocative' function of the texts.

Table 1 General statistics

Text File	OVERALL	T.TXT	N.TXT
Bytes	37,909	20,850	17,059
Tokens	5,727	3,143	2,584
Types	1,501	931	889
Type/Token Ratio	26.21	29.62	34.40
Ave. Word Length	5.42	5.45	5.43
Sentences	241	121	120
Sent. length	23.76	25.98	21.53

(T.TXT=translated texts; N.TXT=non-translated texts)

4.2 Frequency data and Concordance

The high-frequency lexical words usually characterize the aboutness of a text or corpus, and the top frequent lexical words in the two corpora are shown in Table 2. The high-frequency words in both collections are almost the same except for a slight difference in the choice of *service* vs. *serve*, and *attorney* vs. *lawyer*. The non-translated texts tend to make a greater use of *attorney* than the translated ones, which, as we find, is mainly due to the greater proportion of U.S. texts in our collection. Wordlists like this fail to tell about the differences in the informative function between the two collections—it seems both focus on the same information, and I resort to the Keyword

operation. In the Wordsmith Tools program (Scott, 1999) keywords are defined as words whose frequency is unusually high in comparison with some norm. Keywords are identified by comparing two word lists which have been created using the program's wordlist tool, and the loglikelihood statistical test is used to calculate keyness. Keywords which are statistically significantly more frequent in the study corpus than in the reference corpus are called 'positive key words', and statistically significantly infrequent ones 'negative key words'.

Table 2 Top-frequency lexical words

	T. txt			N.txt		
N	Word	Freq.	%	Word	Freq.	%
1	Firm(s)	77	2.45	Client (s)	53	2.05
2	Law	73	2.32	Attorney /Lawyer(s)	36	1.39
3	Legal	53	1.69	Firm(s)	33	1.28
4	Lawyer/ Attorney(s)	36	1.15	Law	28	1.08
5	Service	30	0.95	Legal	28	1.08
6	Client(s)	22	0.70	Serve	24	0.93

In my corpus study, the two subcorpora are compared with each other, i.e. the non-translated collection acts as a reference corpus, thus representing the norm, for the translated one. Such a comparison reveals those words in terms of which the two corpora quantitatively differ the most from each other. See Table 3:

Table 3 Keywords

N		WORD	Freq	T.TXT %	Freq	N.TXT %	Keyness	P
positive	1	China	19	0.60	0		22.9	0.000002
	2	University(ies)	18	0.57	0		21.6	0.000003
	3	Firm(s)	70	2.23	23	0.89	16.8	0.000041
	4	Foreign	13	0.41	0		15.6	0.000077
negative	5	Client(s)	22	0.70	53	2.05	20.3	0.000007
	6	We	13	0.41	41	1.59	21.4	0.000004
	7	Our	19	0.60	92	3.56	68.9	0.000000

I eliminate those proper words which denote the location of the firms, like *Beijing*, *Louisiana*, etc. According to the result, words significantly more frequent in the translated texts are *China*, *university(ies)*, *firm(s)* and *foreign*. The greater use of the word *foreign* is mainly affected by a change in the potential addressee of the advertisements, since in most cases it stresses the firms' international service and therefore does not represent any cultural differences in the realization of the textual function. Concordance of other keywords, however, has revealed them to be significant factors in signaling a difference in informative and vocative functions between the two collections.

First, there seems to be a contrast in the type of information emphasized in the two types of texts. In the non-translated texts, the two sides of the communicative action—the addresser (here realized as *we*, *our*) and the addressee (*clients*)—seem to be foregrounded. In contrast to it, the translated text tend to stress a close relationship with the government and educational institutions, as can be seen through concordance of the words *China* and *university*.

The word *China*'s frequent collocation with government agencies and well-known

universities helps to create a better image of the enterprises, showing the addressers' intention to establish authority by stressing official authorization or permission, their close relation with the government, and their social influence, as can be seen in the following examples:

xxx Firm is a partnership law firm registered in Beijing with the approval from the direct government agency of **China**....

xxx is able to maintain a close relationship with **China's** legislative department and government.

To addressees in the target culture, a close relationship with the government has little to do with the efficiency and competitiveness of the enterprise. Furthermore, approval from the government agencies is a necessary condition for the legal operation of an enterprise, that is to say, the information is already presupposed and therefore redundant in the translations. According to Beaugrande & Dressler (1981), this type of information which is already known to the readers should be 'first order of informativity', of less significance and even boring to the readers, thus indirectly affecting the vocative function as well as the acceptability of the translations.

There are mainly two cases of use for the word *university*, one for describing the education of the staff, the other for illustrating the enterprises' enthusiasm in supporting education. For instances,

The head partners and lawyers graduated from the top domestic **universities** such as Tsinghua **University**, Peking **University**,

xxx firm donated 100,000 yuan to set up Scholarships in The East China **University** of Politics and Law.

While enterprises' participation in social welfare work serves to show their strength, competitiveness and public spirit, and promotes their image and reliability to some degree, the

frequent mention of China's universities may add to a text's informativity and can be a burden on the comprehension process, since the university names might not be part of the target text receivers' cognitive context and can not activate their schemata. The following passage is a typical example of this phenomenon:

In 2003, xxx Firm donated 1,200,000 yuan to set up Teaching Awards and Scholarships in its name on its 10 anniversary in 6 law schools in Peking **University**, Qinghua **University**, China People's **University**, China **University** of Political Science and Law, Jilin **University** and Heilongjiang **University** respectively. In 2004, the firm donated 100,000 yuan to set up Teaching Awards and Scholarships in the First middle **school**, Hailun city, Heilongjiang province. In 2005, the firm donated 100,000 yuan to set up Scholarships in East China **University** of Politics and Law....

Readers from different cultural backgrounds are usually constrained by their knowledge structure and culture differences in their comprehension of textual information, which might be significant for the source text readers and completely the opposite for the target readers. In non-literary translation, the translator often needs to determine which kind of text will best suit the target audience, and whether a text needs to be re-targeted for a new audience, i.e., adjusted in order to meet the new users' information needs and cultural expectations (Varantola, 2003). In the case of the above passage, the translator might employ the technique of omission so as to make it more easily understandable to the target readers, with almost the same effect:

xxx Firm has set up Teaching Awards and Scholarships in middle schools as well as law schools of many well-known universities in support of education in China.

Apart from the informative function, keyword analysis also indicates a difference in the vocative function between the two types of texts, mainly as a result of the different ways of use of

referential terms. Referential terms are often good markers of focalization of a text, which might be internal or external.

Concordance shows that 61 out of 70 occurrences of *firm* in translated texts collocate with the determiner *the* and direct names to refer to the enterprise. This typical way of referring to the enterprises reflects the objective external point of view of the advertiser. That is, the advertiser's mentioning of an enterprise by its direct name or by the third person referential term *the firm* introduces an objective description about it, thus making the readers see things from the point of view of an outsider. The advertisement seems to be presented from some authority who is trying to convince the readers of its objectivity and reliability, resulting, however, in an increase in the psychological distance between the reader and the advertiser.

In contrast, there are only 23 occurrences of *firm* in the non-translated texts, among which only 11 are used as direct reference to the enterprise, and 6 collocate with *our* as self-references. First-person reference is typically employed in translated texts, as manifested by the abundant use of *our* and *we*, signaling internal focalization from the advertisers' point of view. The advertisers seem to be merged with the enterprises, and the 26 cases of collocation of *our* with *client(s)* imply the close relationship between the enterprises and their customers. This is in strong contrast with the non-translated texts in which *client(s)* mostly go with *its* or *the*. Compare the following instances,

xxx Law Firm contrives to provide quality and tailored services to **its clients** with team forces. (T.TXT)

The Firm compensates **the clients** for the loss if **their** interests were damaged due to **its** lawyer's serious faults. (T.TXT)

We also provide **our** clients with a range of private client services... (N.TXT)

For additional information on ... please contact **us** and **we** will be pleased to assist you.

(N.TXT)

Use of personal pronoun *we* and *our* implies the existence of the 'you', which add to the dialogic effect of the texts. In addition, this massive use of *we* as well as *our* helps to transfer the reader subliminally to the internal point of view of the advertiser, resulting in the former's identification with the latter and therefore a decrease in the psychological distance between the two.

The prime principle determining any translation process is the purpose of the overall translational action (Nord, 1997: 27). Although identical in their ultimate purpose (to appeal to text receivers), the two types of texts demonstrate different perspectives in fulfilling this task, which is mainly due to the differing conventions of the same genre in Chinese and English contexts. Since genre conventions are mostly culture-specific, they play an important role in functional translation. If a target text is to be acceptable as representative of a target-culture genre, the translator has to be familiar with the conventions that the target text is to conform to (ibid: 54). Pragmatic translation requires the translator to determine whether a text needs to be re-targeted for a new audience in order to achieve functional equivalence.

5. Conclusions

In this paper we set out to demonstrate how we might use a comparable corpus as an aid to the assessment of Chinese-English pragmatic translations. Small specialized corpora resolve issues pertinent to genre-specific languages. The corpus does not need to be sophisticated in terms of syntactic or semantic tagging. In fact it can be structurally very simple, text-only corpora, but they can nevertheless be very useful in the actual decision-making process in translation.

Translators may compile and use such a corpus to look for stylistic information (through general statistics) and special field term (through Wordlist), and corpus tools like Keyword analysis might also help in translation evaluation, that is, they help the translators to find out how their own translations differ from non-translations in informative and vocation functions. If translators do not find solutions to these type of information needs, a mismatch may arise between translators' competence and their performance (see also Varantola, 2003). On the basis of the statistics gathered with the help of corpus tools, it is possible to probe into the possible causes for these problems from the extra-linguistic factors, and improvement can be made if such factors are taken into serious consideration and constructive measures are adopted.

Compared with the traditional way of mainly subjective assessment, a computer-aided approach has very obvious advantages. As the World Wide Web provides practically unlimited access to electronic texts that can be used in compiling individual disposable corpora for translation, there is no longer any need for the concern about cost efficiency. The knowledge of how to compile and use corpora should be an essential part of modern translational competence and should therefore be dealt with in the training of prospective professional translators.

References

- Baker, M. (1995) 'Corpora in Translation Studies: An Overview and Some Suggestions for Future Research'. *Target* 7 (2): 223-243.
- Baker, M. (1996) 'Corpus-based translation studies: The challenges that lie ahead', in H.Somers (ed.) *Terminology, LSP and Translation: Studies in language engineering, in honour of Juan C.Sager*, 175–86. Amsterdam: John Benjamins.
- Beaugrande & Dressler. (1981) *Introduction to Text Linguistics*. London and New York:

Longman.

Bowker, L. (2001) 'Towards a Methodology for a Corpus-Based Approach to Translation Evaluation'. *Meta* 46 (2): 345-364.

Nord, C. (1997) *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester: St. Jerome.

Nord, C. (2007) 'Looking for Help in the Translation Process—The Role of Auxiliary Texts in Translator Training and Translation Practice'. *Chinese Translators Journal* (1): 17-26.

Olohan, M. (2004) *Introducing Corpora in Translation Studies*. London & New York: Routledge.

Scott, M. (1999) *Wordsmith Tools*. Version 3.0. Oxford: Oxford University Press.

Varantola, K. (2003) 'Translators and Disposable Corpora'. In F. Zanettin, S. Bernardini and D. Stewart (eds.) *Corpora in Translator Education*, 55-70. Manchester: St. Jerome Publishing.

贾文波 (2004) *应用翻译功能论*. 北京:中国对外翻译出版公司.