

Corpus-Based Approaches to Specialized Translation: a New Framework?

*Ilhem Bezzaoucha,
Khadidja Merackchi*

**Institute of Translation,
University of Algiers 2**

Abstract:

Corpus linguistics as a research methodology has been put to many different uses across a wide range of fields as varied as social sciences, natural language processing and cognitive linguistics. Breaking from a long-established tradition of manually analysing large collections of compiled texts, corpus linguistics can unravel some unsuspected patterns in the language of a source text and its target rendition. However, the use of large corpora in Translation Studies is relatively recent. The use of corpus-based methodologies in Translation Studies has known wide applications in specialized translation and terminological extraction. This paper highlights the multiple uses of corpus analysis tools in the extraction of multilingual terminology in the translation of specialized texts. When deciding on a translation equivalent for a specific term or phrase, corpora can prove useful in verifying or rejecting decisions based on other tools such as dictionaries and the Internet.

الملخص باللغة العربية:

لقيت لسانيات المتون اهتماما واسعا في شتى المجالات المعرفية كالعلوم الاجتماعية، والمعالجة الآلية للغات واللسانيات الإدراكية. فبعد تخطي العادة القديمة لتحليل اليدوي لكم هائل من النصوص المجتمعة، كشفت لسانيات المتون عن نمطية جديدة تتميز في نص اللغة المنقول منها و ترجمته. ومع ذلك، يبقى استعمال المتون في الدراسات الترجمية حديثا، ولاسيما تطبيقاته في الترجمة المتخصصة واستخراج المصطلحات.

يسعى هذا المقال إلى الكشف عن أهم استعمالات أدوات تحليل المتون في سياق استقصاء التجليات المصطلحية في ترجمة النصوص المتخصصة. ففي سعيه لاختيار المكافئ الأنسب لمصطلح معيّن أو عبارة، يجد المترجم ضالته في المتون التي قد تفنّد أو تواسي قرارات اتخذت بواسطة وسائل أخرى كالقواميس وشبكة المعلومات.

1. Introduction

One of the basic skills required to be a professional translator is textual competence, which implies the ability to produce specialized target texts that adhere to the conventions and norms of a specialized language. Translation trainees instinctively think in terms of specialized dictionaries and glossaries when they want to access specialized equivalent terms.

Corpora can prove to be equally necessary and useful to translators in the selection of accurate terms as dictionaries updates tend to slower. One of the main trends in corpus-based translation studies is to provide professional translators and trainees with real textual patterns that can be used for terminological references.

This contribution puts forward some ideas of corpus exploitation in the fields of specialized translation and LSP training.

2. Corpora: definitions and usage

According to Bowker and Pearson (2002: 9-10), a corpus (the plural is “corpora”) can be defined as “a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria.” Selection criteria can be as varied as the purposes for which the corpus is built and include size, sampling, population and representativeness among other things (Mc Enery and Wilson, 2001).

One of the most famous English monolingual corpora which had the aim of being broadly representative of British English in the nineties is the British National Corpus, BNC, (<http://corpus.byu.edu/bnc>). BNC is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written. The BNC gathers a large variety of genres of writing, nearly 88 million words about world affairs, social science, leisure, commerce and finance and even

fiction. Just over 10 million words of speech were recorded and transcribed in a consuming timeendeavour. This monolingual corpus can be used by translators to retrieve information about word usage in a rapid and reliable way. A quick search of the word “specialized” in the BNC, for instance, reveals the many varied contexts that it occurs in. This procedure can prove to be of great use when a researcher (translator or language learner) wants to swiftly get the gist of how a specific word relates to other words in a set of different contexts. Corpus analysis will thus reveal important contextual elements such as colligations and collocations (see Figure 1).

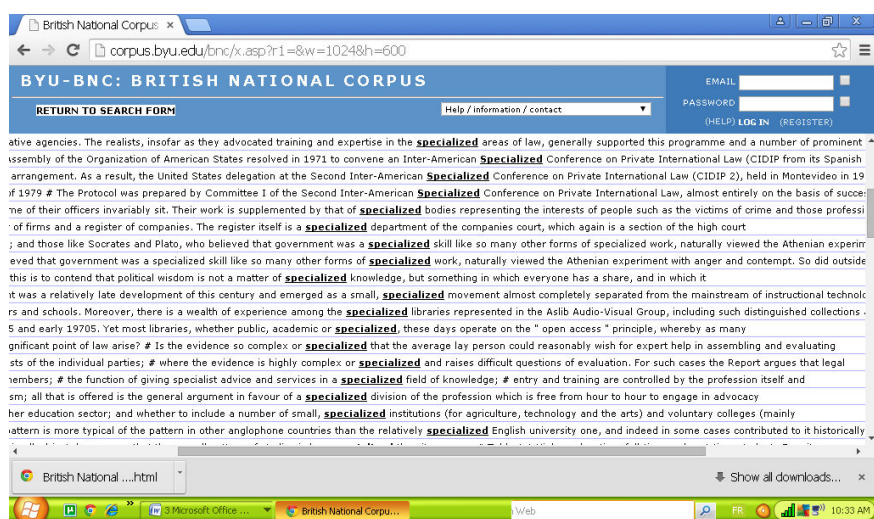


Figure1.BNC interface showing the concordance of the word “specialized”

3. Corpus-based translation studies

With the advent of computers, novel approaches emerged in various scientific fields including translation studies. Large compilations of authentic texts were stored for research purposes. It has been two decades now since Baker (1995) put forward a typology for corpora in translation research and pedagogy: comparable corpora, multilingual corpora and parallel corpora.

As an additional resource for translators, corpora potentially provide readily-made solutions to problems that arise during the translational process, ranging from clarifying terminology, to understanding register or identifying meanings of words in contexts (Pearson, 1998).

Thus, in his translational quest, a translator or a translator trainee can draw from various types of corpora:

- Monolingual corpora: most typical corpora used by linguists, they involve one language: BNC, COCA.
- Monolingual comparable corpora: they involve a corpus of translations and a corpus of texts created spontaneously in the same language: Translational English corpus at the University of Manchester.
- Bilingual comparable corpora: they contain spontaneously created texts in two different languages: Bononia Legal Corpus.
- Parallel corpora: they contain source texts aligned with their translations: CRATER. Examining parallel corpora allows translators to observe the structure, terminology and phraseology typical of genre conventions.

4. Corpus-based Approach-Applications in Specialized Translation

Corpus-based applications are numerous and include various aspects of specialized translation. In recent years, there have been many recommendations by researchers in the field of translation studies to introduce corpora in translation curricula. For example:

In a pilot study, Bowker (1998) found that students using a specialized corpus in the target language did a better job in respect to correct term choice and idiomaticity than a matched group relying solely on conventional resources i.e. bilingual dictionaries, concluding that:

“The general trend shows that for the categories of subject comprehension, term choice, and non-idiomatic expression,

students using the corpus made significantly fewer errors than students using conventional resources. For the remaining two categories, namely grammatical errors and register, students using the corpus performed only marginally better than students using conventional resources. (Bowker, 1998: 46-47).

While Bowler highlights the benefits of the use of monolingual corpora in specialized translation and its advantage for translation trainees, Laviosa (2002: 102-104) reports on how bilingual comparable corpora and parallel corpora can be used in translator training by means of comparison of stylistic features in particular text genre. She also highlights that corpora can also be of benefit not only for translation trainees but also for professionals of translation (Laviosa: 2002: 107).

Another innovative component came to stress the need of using corpora is the constant exposure of practicing translators to the multiple benefits of computer-assisted tools to enhance both the quality and productivity of their work.

“The methodology involved in making corpora and extracting terminology should be part of the teaching curriculum-not optional extra-particularly given the pressure on translators to use computer-assisted tools.” (Maia, 2003: 52)

Referring to the advantages of specialized corpora, Zenattin (1997) states that using specialized corpora enable translators to see how words are naturally used in the target language by real practitioners, and this will enhance the ability of translators to produce fluent and natural target texts.

Another benefit of using specialized corpora in translation is that of acquiring relevant information about the co-occurrence patterns (colligations and collocations) in specialized discourse. Since terms are highly subject field specific, it is possible to acquire a deeper insight of their usage based on their immediate linguistic environment.

Translators regularly consult printed or on-line parallel texts in their language combination in order to search for appropriate terminology, phraseology or check idiomatic usage. Nowadays, translators can consult large quantities of texts far more swiftly if they use corpus analysis tools.

5. Corpus analysis tools in translation studies

In order to process corpora, researchers need corpus analysis tools that help investigate and manipulate information in a corpus. These tools are: Word-frequency lists, type-token ratio, keyword lists, collocations, clusters, concordancers and text aligners. This section focuses on concordancers and text aligners as the most explored features in specialized translation and also because most concordancing softwares do other tasks such as generating keyword frequency lists.

“Concordancing is a means of accessing a corpus of text to show how any given word or phrase in the text is used in the immediate contexts in which it appears. By grouping the uses of a particular word or phrase on the computer screen or in printed form, the concordancer shows the patterns in which the given word or phrase is typically used.”(Flowerdew, 1996:97)

Wordsmith tools version 4 (Smith 2004) and Antconc (Anthony 2014) are among the available and known concordancers. There are also online embedded concordancing programs that allow researchers to get a brief overview of how certain words mingle together.

Corpus analysis programs also offer possibilities to researchers to rapidly compile their own specialized corpora to get a better feel of the usage of certain terms in more specialized contexts. It is a time consuming process to compile a corpus, and many practicing translators resort to already existing parallel corpora to avoid copyright issues.

6. What is a Star then?

We attempted to investigate the term ‘star’ and its various co-occurrences in the BNC and another parallel English-French corpus supported by bilingual concordance called “tradooIT”. Most concordance lines in the BNC appear to be irrelevant as most contexts show examples of ‘star’ being used in other fields such showbiz and social settings. This excessive ‘noise’ as called by Bowker and Pearson (2002) is due to the fact that ‘star’ is a polysemous word. The BNC has, despite its relative large size, limitations as a translation aid if texts deal with specialized knowledge.

TradooIT, a bilingual concordance, gathers a wealth of corpora from many subject-fields. It offers searches in many corpora and several terms banks. Its terminology module finds terms and directs the researcher to the appropriate sites to get further information. The term “star’ for example can be isolated in more specialized corpora dealing with astronomy, tourism or even military dress code, narrowing down the translator’s choices.

7. Conclusion

Specialized translators have relied on parallel and comparable texts (in hard copy) when dealing with terminology-related problems, hence the relative novelty of this technique. Digitized corpora, however, offer the professional translator the opportunity to do the exact same thing at the touch of a button thanks to rapidly developing softwares.

The example used to illustrate terminology extraction suggests alternatives ways for solving terminological problems involving readily available solutions found by other professional translators.

The observation of translation corpora should thus enable translators to track back, through the comparison of source and target texts, analysis and inference, the translators’ choices in rewording and writing that gave to the translated text the form it has (Ballard, 2010: 439).

References

- Ballard, Michel. (2010). "Unit of translation". In *Handbook of Translation Studies*. Volume 1, Gambier, Yves and Luc van Doorslaer; p. 337-340.
- Baker, Mona (1995) 'Corpora in Translation Studies. An Overview and Suggestions for Future Research', *Target* 7(2): 223-43.
- Bowker, Lynne/Pearson, Jennifer (2002), *Working with Specialized Language. A Practical Guide to Using Corpora*. London: Routledge.
- Flowerdew, John, 1996, 'Concordancing in Language Learning', in Martha Pennington (ed.) *The Power of CALL* (Athelstan), pp. 97-113.
- Laviosa, Sara. (2002). *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam – New York, NY: Rodopi B.V
- Maia, Belinda (2003). "Some languages are more equal than others. Training translators in terminology and information retrieval using comparable and parallel corpora". In Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds.) *Corpora in Translator Education*. Manchester: St Jerome 43-53.
- McEnery and Wilson (2001, 2nd ed.) *Corpus linguistics*. Edinburgh Textbook in Empirical Linguistics
- Pearson, J. (1998), *Terms in Context*, Amsterdam, John Benjamins Publishing Company.