

## **CORPORA AND CORPUS TECHNOLOGY FOR TRANSLATION PURPOSES IN PROFESSIONAL AND ACADEMIC ENVIRONMENTS. MAJOR ACHIEVEMENTS AND NEW PERSPECTIVES**

Cécile Frérot\*  
University of Grenoble

**Abstract:** The “use” of corpora and concordancers in translation teaching has grown increasingly attractive since the mid1990s’ with an abundant literature advocating their use and promoting their benefits in the translation classroom. In translator training, efforts are being made to incorporate the use of corpora and concordancers in masters’ programmes and to offer specific modules on corpora for translation as the use of translation memory (TM) systems within Computer-Aided Translation (CAT) courses still dominates. In the translation profession, while TM systems are part of the everyday working environment, the same cannot be said of corpora and concordancers even though the most recent surveys show that professional translators would like to learn more about the potential of corpora for translation. Overall, the “usefulness” of corpora and corpus technology at the different stages of the translation process remains poorly documented in translation but a growing number of empirical studies has started to show concern as it has now become of paramount importance to assess the extent to which corpora are of added value for translation quality in both professional and academic environments

**Keywords:** Corpus. Teaching. Translation Technology. Translator Training. Professional Translator.

---

\* Cécile Frérot: PhD in Natural Language Processing from the University of Toulouse le Mirail (2005). MAS in Linguistics from the University of Paris Diderot (2001). MA in Translation and Language Engineering from the University of Paris Diderot (2000). BA in Translation and Terminology (ISIT, Paris) (1999). Lecturer and researcher in the Department of Applied Languages of the University of Grenoble Alpes. Grenoble, France. E-mail: [cecile.frerot@univ-grenoble-alpes.fr](mailto:cecile.frerot@univ-grenoble-alpes.fr)



Esta obra utiliza uma licença Creative Commons CC BY:  
<https://creativecommons.org/licenses/by/4.0/>

## **CORPUS Y TECNOLOGÍA DE CORPUS PARA LA TRADUCCIÓN EN ENTORNOS ACADÉMICOS Y PROFESIONALES. PRINCIPALES LOGROS Y NUEVAS PERSPECTIVAS**

**Resumen:** Desde mediados de los 90 el “uso” de corpus y programas de concordancias se ha vuelto cada vez más atractivo en la enseñanza de la traducción, de lo que da fe un abundante volumen de publicaciones que apuestan por ello y promueven sus beneficios en el aula. En la formación de traductores se están realizando esfuerzos para incorporar el uso de corpus y programas de concordancias en programas de máster y ofrecer módulos específicos sobre uso de corpus en traducción, si bien aún domina el uso de memorias de traducción (MT) en los cursos de Traducción Asistida por ordenador (TAO). En el mundo profesional de la traducción, mientras que las MT son parte del entorno de trabajo habitual, no se puede afirmar lo mismo de los corpus y los programas de concordancias, a pesar de que los últimos estudios muestran que a los traductores les gustaría saber más sobre el potencial de los corpus para su trabajo. En general, la “utilidad” de los corpus y la tecnología de corpus aún no está bien documentada en el campo de la traducción, si bien existe un creciente número de estudios empíricos en los que se ha empezado a mostrar interés dado que se considera de vital importancia evaluar el valor añadido que aportan los corpus a la calidad de la traducción tanto en el ámbito profesional como académico.

**Palabras clave:** Corpus. Enseñanza. Tecnologías de la traducción. Formación de traductores. Traductor profesional.

### **1. Introduction**

The use of corpora and corpus technology for translation purposes has been on the agenda of teachers in applied translation studies since the mid1990s'. From the teacher's perspective, introducing corpora and concordancers in the classroom has been highly encouraged in order to raise student awareness of language differences, provide student with authentic linguistic material and enhance their translations by providing accurate and idiomatic words and phrases unlikely to be found in more traditional resources (e.g. Bowker and

Pearson, 2002; Zanettin, 2002; Varantola, 2003; Bernardini, 2006).

From the trainer's perspective involved in training future translators, the use of TM systems has become fully integrated in translator training programmes as TM systems are now considered a pre-requisite by Language Service Providers. While Krüger (2012) reminds us that translation didactics has now acknowledged that corpus use forms part of wider translation competence itself that should not be a mere additional qualification to be acquired independently of "pure" translation competence (Rodríguez-Inés, 2009), the use of corpus processing tools - namely concordancers - to search through corpora as part of a translation resource as such has been little applied in translator training institutions (Kübler, 2011, Frankenberg-Garcia, 2015). Overall, the benefits of corpus use as a "performance-enhancing" tool (Varantola, 2003) contributed to the design of translation competence models, the best-known of which are Pacte (Pacte, 2003) and EMT (Expert Group, 2009). While both models focused on the technological side of corpora when they were first devised, they have now regarded corpora as translation aids whose competence features the "ability to use electronic corpora adequately to solve translation problems in an adequate manner" (Rodríguez-Inés, 2009, 136). And at the European Master's in Translation (henceforth EMT<sup>1</sup>) meeting held in March 2015 in Brussels, members of the Working Group on Tools and Translation Technologies featured the use of corpora as a translation resource in training and professional contexts among the most salient themes to be dealt with in the near future. Actually, the use of TM systems dominates in the professional translator's environment and the role of corpora and concordancers as a translation resource needs to be disseminated in the translation profession as evidenced by all recent surveys on the use of corpora by professional translators.

Overall, going beyond the "use" of corpora and corpus technology and investigating their "usefulness" on translation

---

<sup>1</sup> More information can be found at [http://ec.europa.eu/dgs/translation/programmes/emt/index\\_en.htm](http://ec.europa.eu/dgs/translation/programmes/emt/index_en.htm) (last accessed 18 September 2015).

quality at the different stages of the translation process has now become of paramount importance to assess the extent to which corpora are of added value for translation quality in both professional and academic environments. Therefore, in our paper, we aim at providing a panorama on the “use” of corpora and corpus technology in the academic environment (section 2.) that encompasses both applied corpus-based translation studies and translator training (section 2.1. and 2.2. respectively) as well as uncovering the use of such tools in the professional environment (section 3.). This will enable us to bring new insights into the directions taken by the most recent research; in particular, we will put the emphasis on the increasing concern related to the “usefulness” of corpora and corpus processing tools on translation quality and we will address a number of related issues based on an experiment recently carried out with third-year students specializing in translation (section 4.).

## **2. The Use of Corpora and Corpus Technology in the Academic Environment**

### **2.1. Corpora and Concordancers for Translation Teaching Purposes**

As stated by Laviosa (2010), the introduction of corpora in Translation Studies was put forward by Mona Baker (Baker 1995) in her seminal article entitled *Corpus linguistics and Translation Studies: Implications and Applications*. Since then, the use of corpora has grown increasingly attractive in corpus-based translation studies as well as in applied corpus-based translation studies. While the former relates to a corpus-based methodology for identifying the distinctive features of the language of translation and is aimed at understanding “the specific constraints, pressures, and motivations that influence the act of translating and underlies its unique language” (Baker, 1998, 480), the latter is of particular interest to us as our main focus in this section is to consider the

use of corpora and concordancers for translation teaching. Broadly considered, there are two major complementary approaches to using corpora and corpus technology for translation teaching known as “Corpus use for learning to translate” and “Learning corpus use to translate”. To put it as simply as Beeby et al. (2009) in their Introduction to *Corpus Use and Translating* “the first part, *Corpus use for learning to translate* will give ideas to teachers who want to prepare learning materials and tasks using corpora. The second part, *Learning corpus use to translate* is about helping students to become autonomous users of corpora as part of their translation competence” (*ibid.* 1.). More specifically, in the former approach, corpora are provided by teachers who design corpus-based translation-related tasks so that students focus on a particular translation issue and analyze a given set of preselected data.

In the “Learning corpus use to translate” approach, students are mainly taught how to build corpora from scratch as “this approach does not primarily focus on the immediate corpus-use related aspects but instead on the various translation-related issues of corpus compilation, for example, corpus design, search strategies, assessment of potential corpus sources, assessment of the adequacy and relevancy of corpus texts, general software literacy” (Krüger, 2012, 509). In this Do-It-Yourself (henceforth DIY) approach which involves building one’s own corpus - termed DIY or disposable or even ad hoc corpora - such resources are used to help students solve specific translation problems, for instance, find the most accurate words and phrases. As described in Varantola (2003), using disposable corpora help students gain reassurance for their strategic decisions and lexical choices which the author considers a major benefit. As the author puts it: “we can regard them [disposable, ad hoc corpora] as performance-enhancing tools in translation or, more precisely, as decision-making tools for lexical and textual knowledge management in translation.” (*ibid.*, 59). Whichever approach is considered in the classroom, corpora are resources that are meant to provide students with translation solutions that cannot be found - or at least hardly be found - in other

electronic resources such as dictionaries, and as such, corpora and corpus interrogation tools serve as “documentation tools” (Marco Borillo and van Lawick, 2009).

The use of corpora and concordancers for translation teaching purposes is widely documented and the abundant literature stresses the major benefits that can be gained from using different types of corpora in the translation classroom. Those corpora fall into two main categories: *i*) comparable corpora - commonly defined as “a collection of texts composed independently in the respective languages and put together on the basis of similarity of content, domain and communicative function.” (Zanettin 1998: 614), and *ii*) parallel corpora, defined as having “components in two or more languages, consisting of original texts and their translations” (Aston 1999: 290). Comparable corpora have been used quite extensively in the classroom. Overall, they have been praised as enhancing the understanding of the source language text and target language production by providing information missing from dictionaries (Zanettin, 1998), helping students gain insights into the languages and cultures involved (Zanettin, 2001) and promoting language awareness in students’ translations regarding context, text type, register and idiom (Bowker, 1999). Other studies of interest on comparable corpora can be found in Pearson (1996; 2000), Kübler (2003), Maia (2003), Varantola (2003), Wilkinson (2005) and Gauton (2008). As for parallel corpora, they have been said to be an important asset in translation teaching (Olohan, 2004) especially compared with bilingual dictionaries (Williams, 1996) as well as complementary to comparable corpora (Pearson, 2003) but the literature still remains not as abundant as the one on comparable corpora. And to date, most studies have put the emphasis on designing tools to investigate parallel corpora, some of which have clearly stressed the pitfalls of dictionaries and the value of parallel corpora as terminological resources (Peters and Picchi, 1998). Nonetheless, a number of valuable studies are worth being mentioned among which are Pearson (2003), Laviosa (2006) and Fernandes (2007).

## **2.2. Translation Memory Systems and Corpora in Translator Training**

While corpora and concordancers are highly favoured in corpus-based translation-related activities, TM systems in particular are used to train future translators. This type of technology has become increasingly popular in the translation industry over the last fifteen years (Bowker, 2002, Bowker and Barlow, 2008, Garcia, 2009) and is now considered a pre-requisite by Language Service Providers as recently shown by the conclusions drawn from the discussions of the working groups involved in the European Graduate Placement Scheme (henceforth, EGPS). As a matter of fact, a survey carried out among 26 recent graduates<sup>2</sup> in translation at the University of Grenoble Alpes shows that 92% use at least one TM system in their working environment (Frérot and Karagouch, to appear).

In parallel to their primacy on the translation market, TM systems have gained such considerable importance in translator training that it can be claimed that they are now fully integrated in masters' programmes in translation. This is evidenced by the Optimale survey as well by a survey we recently carried out among French universities. The Optimale survey on translation technologies and tools teaching has been part of an overall project aimed at "enhancing the visibility and relevance of professional translator education and training in Europe (...) and providing Master's degree level programmes that equip graduates with a thorough knowledge of the translation industry, professions and processes, and with a range of competences that are relevant to professional requirements in this area"<sup>3</sup>. Its most significant result shows that over 90% of the European masters' programmes

---

<sup>2</sup> The 26 respondents to the survey graduated in 2013 and 2014.

<sup>3</sup> p. 3 of a progress report which can be found at [http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/40/Public%20part\\_report\\_2010\\_OPTIMALE%204018-001-001.pdf](http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/40/Public%20part_report_2010_OPTIMALE%204018-001-001.pdf) (last accessed 25 September 2015)

in translation surveyed<sup>4</sup> train their students on how to use TM systems while the French survey amounts to 100%. Nonetheless, it should be stressed that masters' programmes offer core courses on TM technology with an emphasis on technical know-how rather translation competence, which has been criticized (Chung-ling, 2006; Sauron, 2007; Kenny, 2007). Recent research advocates the use of TM systems in the translation classroom itself for students' benefits (Bowker et al., 2008; Frérot and Karagouch, to appear) as students give major importance to being trained in real working conditions and "translate as real professionals"<sup>5</sup>. In this regard, one challenge that lies ahead in the near future is "bring the workplace into the curriculum" (Kiraly, 2015)<sup>6</sup> by integrating authentic translation projects in the classroom and/or designing project-based translation classes as well as fostering the existence of junior companies as is already the case at the University of Grenoble Alpes<sup>7</sup>. Intensive technical translation sessions as those offered by the University of Rennes where students simulate a real-time translation project<sup>8</sup> or experiments such as the one recently conducted at the University of Grenoble Alpes - where 2<sup>nd</sup> year MA students have been in charge of a real translation project from the translation order to the translation delivery (Lavault and Frérot, under submission) - are meant to "make the curriculum relevant for the workplace"<sup>9</sup> and provide

---

<sup>4</sup> 29 EMT programmes and 40 Optimale programmes in translation responded to the survey.

<sup>5</sup> As strongly suggested by the graduates who responded to our survey.

<sup>6</sup> Personal notes from an oral presentation at the EGPS meeting in Barcelona, February 2015.

<sup>7</sup> Atlas is the junior company ran by 2nd year MA translation students in Grenoble. Detailed information can be found at [http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/15/WP5.3%20Case%20study\\_student%20company\\_Grenoble\\_EN.pdf](http://www.ressources.univ-rennes2.fr/service-relations-internationales/optimale/attachments/article/15/WP5.3%20Case%20study_student%20company_Grenoble_EN.pdf) (last accessed 30 September 2015).

<sup>8</sup> <http://www.sites.univ-rennes2.fr/lea/cftrtr/?q=fr/node/57#tradutech> (last accessed 25 September 2015).

<sup>9</sup> Personal notes from a presentation by Don Kiraly at the EGPS meeting in Barcelona, February 2015.



new insights into how better match masters' programmes with the requirements of the translation market.

While the integration of TM systems in masters' programmes in translation can be regarded as achieved, the question arises as to the extent to which the use of corpora and concordancers within corpus-based translation-related tasks is being performed. Kübler (2011) points out that there are not many translator training institutions teaching novices how to use corpora. This is in line with Frankenberg-Garcia (2015) who states that "a brief look at the current 2014 programme descriptions of MA in Translation programmes offered in fourteen different UK universities shows that less than a handful of these institutions offer specific modules on corpora for translation" and the author adds that the situation is likely to be similar in other countries. Actually, the survey carried out among French universities shows that around 75% claim to provide corpus-related training but a closer look at the results shows a wide variety of issues addressed in the training ranging from term extraction, natural language processing, electronic tools and web search, DIY approach for translation; the Optimale surveys found that 60% focused on corpus construction but did not offer it as a major course. Overall, these surveys call for (re)defining what a corpus is and reinforcing the role of corpora among trainers and trainees as already claimed by Bernardini and Castagnoli (2008), who underlined that substantial efforts had to be made to raise both student and teacher awareness of the potential of corpus work in the classroom. Recent conclusions drawn from members of the Working Group on Tools and Translation Technologies at the EMT meeting held in March 2015 in Brussels agreed that the role of corpora among trainers and trainees had been "a matter of improvisation for a certain period of time<sup>10</sup>". One of the major decisions was to re-run the 2012 Optimale survey with updated questions including corpora as well as conduct

---

<sup>10</sup> Personal notes from the EMT meeting in Brussels, March 2015.

a parallel survey of employers given the lack of data in the field with the overall objective of helping masters' programmes match demand from the translation market.

Other directions to be taken are likely to involve the use of large online translation databases (e.g. Linguee, Tradoo) since these are considered as complementary to TM systems based on the survey among recent graduates. In particular, graduates view such resources as a “source of inspiration”, shedding “new light on translation solutions” and providing “new clues”. Integrating such resources will necessarily imply raising student awareness of the limits of such tools in terms of translation quality and assess their usefulness compared with other online resources.

### **3. The Use of Corpora and Corpus Technology in the Professional Environment**

The technology at translator's disposal in his/her working environment includes TM systems and other CAT tools, machine translation tools as well as web-based resources such as dictionaries and termbanks, portals, directories of dictionaries and glossaries, specialized search and meta search engines as well as corpora. While translation technologies have proven to be indispensable for professional translators (Bowker and Corpas Pastor, 2015), not all of them have been used to the same extent and surveys on the use of translation technology by professionals<sup>11</sup> have shown that not all the existing technology varieties were successfully adopted by professionals (Zaretskaya et al., 2015)

Most importantly, not all surveys have integrated the use of corpora by translators - the common ground of all surveys being

---

<sup>11</sup> The most reviewed in the literature are Bowker (2002), Lommel & Ray (2004), Fulford & Granell-Zafra (2005), eCoLoTrain (2006), Lagoudaki (2006), Mellange (2006), TradOnline (2010), Torres Domínguez (2012), Gallego-Hernández (2015) and Picton et al. (2015).

TM systems<sup>12</sup> - and while “in today’s market, the use of technology by translators is no longer a luxury but a necessity if they are to meet rising market demands for the quick delivery of high-quality texts in many languages” (Bowker and Corpas Pastor, 2015), there seems to be little or no demand from the translation market requiring translators to use corpora and concordancers in particular. Actually, there is no particular pressure from the market to train translators to use corpora (Frankenberg-Garcia, 2015) which, in turn, does not encourage the integration of corpora in translator training and may be “counterproductive” for teachers involved in the training of future translators.

However, a number of surveys - though limited - has set out to uncover the use of corpora and corpus technology by professional translators and has shed some interesting light on the professional practice. To begin with, Bowker (2004) carried out a literature survey of publications produced by Canadian translators’ associations and also analyzed a database of job advertisements to “determine how many Canadian employers are seeking candidates who are familiar with corpus-based resources” (*ibid.*, 213). Her major results show that Canadian professional translators are more interested in TM systems than in corpora even though they know that corpora exist. As for the job advertisements for translation-related positions, 62.3% stated some general degree of computer literacy as required and a mere 15.1% were aimed at candidates with knowledge of specialized translation software. But the advertisements never mentioned the word “corpora” as part of their requirements.

A few years later, a survey on the use of corpora by professional translators was conducted within the Mellange project (Mellange 2006) with a total of 740 respondents - the vast majority were from the UK (567) and France (125) and a limited number from Germany (25), Italy (19) and Spain (4). Overall, the data collected brought to light that corpora and concordance use were not common practice

---

<sup>12</sup> The use of CAT tools by professionals falls beyond the scope of our study as our main concern is the use of corpora and corpus technology.

among professional translators and that translators (53.4%) “read” electronic corpora and used search facilities in word processors instead of searching them with software. The survey also showed that professional translators found it time-consuming to build corpora but - and that is of major relevance for the future of corpora - 83.7% had major interest in the potential of corpus use.

More recently, Gallego-Hernández (2015) surveyed the Spanish market and provided data on the use of corpora as translation resources from 526 Spanish translators of various ages and varying amounts of experience. The overall results suggest that integrating skills related to corpus use when designing translation courses combined with the abundant literature on building and exploiting corpora “are beginning to bear fruit, as almost 50% of respondents stated they use corpora ‘sometimes/often/very often’ in their work” (*ibid.*). Another noteworthy result is that concepts related to corpora and how to exploit them may remain poorly defined among professional translators, “or at least have rather diffuse boundaries” (*ibid.*). This is in line with Zanettin (2002) and Olohan (2004) who found that professional translators had little explicit knowledge of corpora. In this respect, Carratalá-Puertas (2015) describes the relationship between corpora and professional translators “as omnipresent as invisible<sup>13</sup>” claiming that professional translators actually resort to corpora as they look for parallel texts using search engines or search previous translations. Picton et al. (2015), conducting a survey in order to assess the translators’ use of corpora in Switzerland based on the gap between the scholars and the translators’ point of view, found that respondents to the survey hardly use corpora as such on a systematic basis and use Google as a mega-corpus (rudimentary use); the survey also found that translators are rather confused when asked to define what a concordancer is and have no clear knowledge of the word “corpus”. So it looks like defining what a corpus is as well as defining the corpus-related concepts should

---

<sup>13</sup> Featured in the abstract book of the conference, p. 28 and can be found at <http://dti.ua.es/en/comenego/iv-cult/programme.html> (last accessed 30 September 2015).

be part of the core programme of any translator training institution. Finally, respondents to Gallego-Hernández's survey among corpus users said they favoured specialized and multilingual corpora over general and multilingual corpora, as well as stable corpora. This echoes to our survey among recent graduates at the University of Grenoble Alpes who claimed that they widely used multilingual large online translation databases (Linguee ranked first in the results) and hardly resorted to DIY corpora. Be it the European Mellange survey or the Spanish one, corpus users acknowledge the benefits of corpora and require specific training, which calls for a greater collaboration between academic and professional circles. Besides, the interest non-users have in the potential corpora offer should greatly motivate translator trainers "who are the real transmitters of this type of methodology" (Gallego-Hernández, 2015) and foster the dissemination of corpus benefits. For instance, Frérot (2011) searched information on professional translators using corpora and concordancers but it proved fruitless as there didn't seem to have any review or report on the use of corpora by translators themselves (in contrast with a number of reviews on CAT tools), nor professional association literature (online forums for instance). A few years later, the picture is quite the same as Frankenberg-Garcia (2015) examined the international translator forums at Proz.com and TranslatorsCafe.com and found that they did not contain any threads about corpora, compared to several daily queries about translation-memory systems and CAT tools in general.

#### **4. The "Usefulness" of Corpora in Translation Quality**

While corpora can serve as tools to evaluate translation quality, for instance to identify the extent to which student translations differ from professional translations (Cappelle et al. 2016<sup>14</sup>), their

---

<sup>14</sup> The authors empirically analyzed the distribution of general linguistic features in student translations compared to professional translations; while their major

usefulness may also be considered with regard to their added value on the translation process itself. One noteworthy study is Lynne Bowker's - back in the late nineties - who investigated the usefulness and impact of corpora on the quality of students' translations. Her pilot study (Bowker, 1998) - quite unique in its kind - aimed to reveal how valuable a specialized monolingual native language corpus could be compared with conventional resources such as dictionaries. Fourteen English native-speaker students, enrolled in an Applied Computational Linguistics programme, translated two French texts<sup>15</sup> into English using conventional resources on the one hand and a corpus together with a concordancer<sup>16</sup> on the other. The results showed that corpora help enhance students' translations by providing information missing from dictionaries, especially regarding term choice and idiomatic expressions.

Recent research is increasingly concerned as it has started to investigate the extent to which corpora are of added value for translation quality. A small but growing number of studies have tackled the usefulness of corpora at an experimental level by collecting data produced by students specializing in translation. Most recently, Kübler et al. (2015), while stressing how poorly documented experimental research was when it came to assessing the usefulness of corpora on the translation process, investigated the efficiency of corpus use for students during terminology processing and LSP translation in the field of earth sciences and found that "besides an obvious and expected positive outcome on the validation and translation of terms (...) there is an interesting positive influence of corpora in the translation process on elements other than those related to terminology, such as collocations

---

results show a number of distinctive features between both groups (lower lexical creativity, more explicitness, longer sentences and higher lexical density were found in student translations), they also show no significant differences in terms of word length, formulaicity or word classes, for instance.

<sup>15</sup> 300 words each, extracted from *Science et vie Micro*.

<sup>16</sup> WordSmith Tools, <http://www.lexically.net/wordsmith/version3/index.html> (last accessed 30 September 2015).

and various genre- and discourse-based features<sup>17</sup>". Heylen and Verplaetse (2015) investigated the usefulness of parallel corpora for medical translation training and analyzed their impact on student performance. In particular, they measured the added value of parallel corpora through the specific use of translation memories versus bilingual concordance search within CAT tools and external sources. While all studies agreed on the added value of corpora on terminology, corpora were also found to be of great help for retrieving idiomatic expressions (Bowker, 1998) and identifying discourse-based features (Kübler, 2015); Heylen and Verplaetse (2015) pinpointed the major benefit of concordance search and external sources compared with translation memories.

Actually, conducting such experimental studies involves considering issues that particularly relate to:

(i) defining among the different stages of the translation process which stage(s) to incorporate in the study i.e. understanding of the subject field, lexical, terminological, phraseological and syntactic accuracy - in relation with conceptual and linguistic information retrieval - as well as stylistic fluency and appropriacy;

(ii) selecting the various corpora to be assessed and to be compared: DIY corpora (monolingual, comparable, parallel corpora), online corpora (for instance the BNC or the COCA for the English language), large online translation databases (e.g. Linguee, Tradoit or Reverso Context) as well as Google search; to our mind, this issue is of major importance as it has an impact on the corpora and corpus processing tools to integrate into the translation classroom based on their efficiency on student translations;

(iii) determining the criteria to consider in order to measure the usefulness of the corpora under study (for instance, which linguistic criteria).

---

<sup>17</sup> The quotation features in the abstract book of the conference, p. 26 and can be found at <http://dti.ua.es/en/comenego/iv-cult/programme.html> (last accessed 30 September 2015).

Conducting such experimental studies also involves devising an adequate methodology at the experimental level that includes determining (i) which instructions give students participating in the experiment - especially in terms of providing feedback on their search and on their use of electronic resources - and (ii) the linguistic items *or* portion of text to evaluate; both options have already been investigated as Kübler (2016) had her students translate 200 words while Heylen and Verplaetse (2015) focused on 10 isolated items (terms such as *abdominal pain*, *adverse reactions* and *hepatitis enzymes*).

We have started addressing the aforementioned issues at the University of Grenoble Alpes in an experiment which sets out to assess the usefulness of corpora for translation, more specifically the usefulness of free online resources integrated in a course on *tools and resources for translators* (term banks, large translation databases and specialized search engines in particular) compared with the Scientext corpus, a free online specialized corpus in biology and medicine<sup>18</sup> which the students participating in the experiment have been trained to use for translation purposes. The experiment consisted in having 16 third-year French students specializing in translation and enrolled in the above mentioned course translate into English 12 linguistic items extracted from a French paper on the diagnosis and management of prosthetic vascular graft infections. The linguistic items were all prepositional phrases French students hardly master to translate as they tend to overuse the “the [Noun1] of [Noun2]” construction as a loan translation (e.g. *qualité de l'image* translated as *quality of the image*) where the “[Noun2][Noun1]” construction (e.g. *image quality*) may be more appropriate (Rossi et al. 2016). Therefore, from a methodological perspective, we

---

<sup>18</sup> The English subcorpus of the Scientext corpus is a 35-million word English corpus made up of papers in Biology and Medicine extracted from BioMed Central. ScienQuest - a web-based concordancer designed for searching Scientext - is freely available at <http://corpora.aiakide.net/scientext18/> (last accessed 30 September 2015).



decided to first identify a given translation problem faced by our students and then assess the extent to which corpora could help them enhance their translations. The overarching aim being to evaluate the added value of a specialized corpus (Scientext) compared with other online available resources, students were instructed to first translate the 12 linguistic items using the Web excluding Scientext (1<sup>st</sup> translation) and then turn to Scientext only in order to enhance their translations, where appropriate (2<sup>nd</sup> translation). In particular, students were asked to indicate which resources they had used for their 1<sup>st</sup> translation and were also encouraged to comment on the difficulties encountered during the whole experiment.

As mentioned earlier, an issue of major relevance relates to determining the criteria to consider in order to measure the usefulness of the corpora under study. While former studies have relied either on different levels of translation acceptability (Delizée, 2010 mentions three levels termed as “accepted”, “to be revised” and “unacceptable” and Heylen and Verplaetse use the “yes/no/partly” criteria for a correct translation) or on an error-annotated corpus<sup>19</sup> (Kübler, 2016), we have used four different criteria, namely compliance with the gold standard reference translation (*i.e.* the English translation of the French paper used in the experiment), acceptability and non-acceptability<sup>20</sup> as well as lack of translation when students provided no translation.

Analyzing the translations consisted in comparing both translations taking into account the four criteria just mentioned. Overall, the analysis shows promising results with an added value of 12% between both translations as using Scientext helped students modify their 1<sup>st</sup> translation (e.g. *treatment duration* / *sedimentation*

---

<sup>19</sup> The student translations have been error-annotated using the Mellange annotation scheme (Mellange 2006).

<sup>20</sup> Based on linguistic criteria used in quality assurance with regards to terminology, meaning, grammar, punctuation and spelling in particular; see Lavault and Allignol (2014) for an overview of the notion of acceptability in professional translation.

*velocity*) which then either complied with the reference translation (*duration of treatment*) or proved to be enhanced (*sedimentation rate*). In contrast, the results also show a significant loss in translation quality as students seem to rely heavily on the Scientext corpus (for instance, *ponctions d'abcès* first translated by *abscess punctures* - the actual reference translation - was changed into *abscess drainage* as no occurrence of the former was found in Scientext). Generally speaking, this “blind faith” in the corpus needs to be tackled in the corpus-based translation classroom where the emphasis should be put on raising student linguistic awareness since using a corpus as a translation resource requires interpreting the data (Kübler, 2016). In this regard, Bernardini (2015) goes as far as suggesting the introduction of a course on corpus linguistics outside the translation classroom in order to help students realize the potential of corpora. Finally, and quite unexpectedly, around 70% of the students provided no 2<sup>nd</sup> translation. While this may be partly due to a lack of time to complete the whole experiment or to technical problems using Scientext, it requires redefining the instructions given to students and ask them to explicitly mention the reason why they may not provide any 2<sup>nd</sup> translation.

## 5. Conclusion

In this paper, we provided a panorama on the “use” of corpora and corpus technology in the academic environment in both applied corpus-based translation studies and translator training and uncovered the use of such tools in the professional environment. We took stock of the major achievements in the field and raised new perspectives; in particular, we addressed new emerging issues as recent research is increasingly concerned with the “usefulness” of corpora and corpus processing tools on the translation process and sets out to investigate the extent to which corpora have an impact on translation quality. Experimental studies such as the one recently carried out in the University of Grenoble Alpes aims at

raising student awareness of the benefits of corpora as they find it difficult to realize the usefulness of a given corpus or DIY corpora otherwise and find it a lot more relevant to turn to Google especially due to time constraints to deliver a translation (Bernardini, 2015).

Therefore, providing both future translators and professional translators with empirical evidence on the value of corpora is very likely to serve both groups and raise their awareness of corpora on translation quality since “for translators to start using corpora, it is important that they realize that corpora have the potential to help them find answers to questions for which there are often no clear answers in dictionaries, glossaries, Google searches and other tools and resources they are accustomed to using.” (Frankenberg-Garcia, 2015). The dissemination of a corpus-based approach in the translator working environment will hopefully bear fruit and one can hope a widespread uptake in professional practice. Overall, the future of the usefulness of corpora in translation will continue to be shaped by empirical studies addressing the thorny problem of translation quality which will undoubtedly contribute to enrich the field and provide answers to issues related to corpora and teaching in particular.

## **Acknowledgments**

I would like to thank Caroline Rossi in particular, who prepared the experiment (see section 4. The “Usefulness” of Corpora in Translation Quality) together with me, provided input throughout the preparation phase on scientific and methodology issues and co-designed with me the grid that was shared with the students.

I would also like to thank Caroline Rossi, who got hold of Pierre-Emmanuel Colle’s two texts, both the published translation and the unpublished original, and was allowed by the author to use them in this experiment. We would also like to thank Pierre-Emmanuel Colle for letting us do so.”

## References

- Aston, G. "Corpus use and learning to translate." *Textus* 12 (1999): 289-314.
- Baker, Mona. "Corpora in translation studies: An overview and suggestions for future research." *Target* 72 (1995): 223-244.
- Baker, Mona. "Réexplorer la langue de la traduction: une approche par corpus." *Meta* 43.4 (1998): 480-485.
- Bédard, C. "Mémoire de traduction cherche traducteur de phrases..." *Traduire* 186 (2000).
- Beeby, A., Rodríguez-Inés, P., & Sánchez-Gijón, P. "Introduction." *Corpus Use and Translating*. Ed. A. Beeby, P. Rodríguez Inés and P. Sánchez-Gijón. Amsterdam: John Benjamins, 2009. 1-8.
- Bernardini, S., Castagnoli, S. "Corpora for translator education and translation practice." *Topics in Language Resources for Translation and Localisation*. Ed. E. Yuste Rodrigo. Amsterdam/Philadelphia: John Benjamins, 2008. 39-55.
- Bernardini, S. "Exploratory learning in the translation/language classroom: Corpora as learning aids." *CULT Conference, 26-29 May*, Alicante, 2015.
- Bowker, L. "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: a Pilot Study." *Meta* 43.4 (1998): 631-651.
- Bowker, L. "Exploiting the potential of corpora for raising language awareness in student translators." *Language Awareness* 8 (1999): 160-73.
- Bowker, L. *Computer-Aided Translation Technology: A Practical Introduction*. University of Ottawa: Ottawa Press, Didactic of Translation Series, 2002.
- Bowker, L. "Corpus Resources for translators: academic luxury or professional necessity?" *TradTerm* 10 (2004): 213-247.

Bowker, L., Pearson, J. *Working with Specialized Text: A Practical Guide to Using Corpora*. UK: Routledge, 2002.

Bowker, L., Barlow, M. "A comparative evaluation of bilingual concordancers and translation memory systems." *Topics for Language Resources in Translation and Localisation*. Ed. E. Yuste Rodrigo. Amsterdam/Philadelphia: John Benjamins, 2008. 1-22.

Bowker, L., McBride, C., Marshman, E. "Getting more than you paid for? Considerations in integrating free and low-cost technologies into translator training programs." *Redit* 1 (2008): 26-47.

Bowker, L., Corpas Pastor, G. "Translation Technology." *Handbook of Computational Linguistics*. Ed. R. Mitkov. Oxford: Oxford University Press, 2015.

Cappelle, B., de Clercq O., de Sutter, G., Loock, R., Plevoets, K. "A corpus-based, statistical approach to translation quality." *Journée d'études Traduction & Qualité 2016: corpus et qualité*. Université Lille 3, 2016.

Carratalá-Puertas, I. "Corpus y traducción profesional, una relación tan omnipresente como invisible." *CULT Conference, 26-29 May*, Alicante, 2015.

Chung-ling, S. "Using Trados's WinAlign Tool to Teach the Translation Equivalence Concept." *Translation journal* 10.2 (2006).

Danielson, P., Ridings, D. "Corpus and Terminology: Software for the Translation Program at Göteborgs Universitet, or Getting Students to do the Work." *Multilingual Corpora in Teaching and Research*. Ed. S.P. Botley, A.M. McEnery, and A. Wilson, 2000. 65-72.

Delizée, A. "Évaluation sommative de la traduction pragmatique en dernière année de formation." *Journée d'études Traduction et Qualité: Méthodologies en matière d'assurance qualité*, 5 February, Université Lille 3, 2010.

Dragsted, B. "Computer-aided Translation as a Distributed Cognitive Task." *Pragmatics & Cognition* 14.2 (2006): 443-464.

eCoLoTrain. Translator training survey – results, 2006.

EMT Expert Group. Competences for professional translators, experts in multilingual and multimedia communication, 2009.

Frankenberg-Garcia, A. “A Peek into What Today’s Language Learners as Researchers Actually Do.” *International Journal of Lexicography* 18.3 (2005): 335-355.

Frankenberg-Garcia, A. “Training translators to use corpora hands-on: challenges and reactions by a group of 13 students at a UK university.” *Corpora* 10.2 (2015).

Fernandes, Lincoln.-P. “On the Use of a Portuguese-English Parallel Corpus of Children’s Fantasy Literature in Translator Education.” *Cadernos de Tradução* 20 (2007): 141-163.

Frérot, C. “Translation Technology in the Classroom: Integrating Bilingual Corpora and Corpus Analysis Tools.” *Monterey Forum on Translator Education*, 7-9 April, Monterey, 2011.

Frérot, C., Karagouch, L. (to appear). “Outils d’aide à la traduction et formations de traducteurs : vers une adéquation des contenus pédagogiques avec la réalité technologique des traducteurs.” *ILCEA* [online], 27 (October 2016).

Fulford, H., Granell-Zafra, J. “Translation and Technology: a Study of UK Freelance Translators.” *The Journal of Specialised Translation* 4 (2005): 2-17.

Gallego-Hernández, D. “The use of Corpora as translation resources: a study based on a survey of Spanish professional translators.” *Perspectives: Studies in Translatology* 23.3 (2015): 375-391.

Garcia, I. “Beyond Translation Memory: Computers and the Professional Translator.” *The Journal of Specialised Translation* 12 (2009): 199-214.

Gauton, R. “The use of corpora in translator training in the African language classroom.” *Topics for Language Resources in Translation and Localisation*. Ed. Yuste Rodrigo. Amsterdam/Philadelphia: John Benjamins, 2008. 89-106.

Granger, S. "The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies." *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Ed. S. Granger, J. Lerot and S. Petch-Tyson. Amsterdam & Atlanta: Rodopi, 2003. 17-29.

Heylen, K., Verplaetse, H. "Parallel Corpora for Medical Translation Training: An Analysis of Impact On Student Performance." *CULT Conference, 26-29 May*, Alicante, 2015.

Johansson, S., Hofland, K. "The English-Norwegian Parallel Corpus: Current Work and New Directions." *Multilingual Corpora in Teaching and Research*. Ed. S.P. Botley, A.M. McEnery, and A. Wilson. 2000. 134-147.

Kenny, D. "Translation memories and parallel corpora: Challenges for the translator trainer." *Across Boundaries: International Perspectives on Translation Studies*. Ed. D. Kenny and K. Ryou. Newcastle: Cambridge Scholars Publishing, 2007. 192-208.

Kiraly, D. European Graduate Placement Scheme (EGPS) event. Barcelona, 5-6 February, 2015.

Krüger, R. "Working with corpora in the translation classroom." *Studies in second language learning and teaching* 4 (2012): 505-525.

Kübler, N. "Corpora and LSP translation." *Corpora in Translator Education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St Jerome, 2003. 25-42.

Kübler, N. "Working with Corpora for Translation Teaching in a French-speaking setting". *New Trends in Corpora and Language Learning*. Ed. A. Frankenberg-Garcia, L. Flowerdew, and G. Aston. London: Bloomsbury, 2011. 62-80.

Kübler, N., Pecman, M., and Volanschi-Mestivier, A. "A study on the efficiency of corpus use for translation students during terminology processing and LSP translation." *CULT Conference, 26-29 May*, Alicante, 2015.

Kübler, N. "Traduction spécialisée: mais où est le terme et comment le traduire?" *Journée d'études Traduction & Qualité 2016: corpus et qualité*, 5 February, Université Lille 3, 2016.

Lagoudaki, E. *Translation memory systems: Enlightening user's perspective*. Translation memories Survey 2006, Imperial College, London, 2006.

Lavault, E., Frérot, C. (under submission). "Raising Ergonomic Awareness of the Workplace in Translator Training Curriculum." *EST Congress 2016, Arrhus University*.

Lavault, E. Allignol, C. "La notion d'acceptabilité en traduction professionnelle : où placer le curseur ?" *ILCEA 19* (2014) (last accessed 21 February 2016). <http://ilcea.revues.org/2455>.

Laviosa, S. "Data-driven learning for translating Anglicisms in business communication." *IEEE Transactions on Professional Communication*. Special Issue on Insights from Corpus Linguistics for Professional Communication. Guest Editor: Thomas Orr, 49.3 (2006): 267-274.

Laviosa, S. "Corpora." *Handbook of Translation Studies*. Y. Gambier, L. van Doorslaer, 1, 2010. 80-86.

Lommel, A., Ray, R. LISA (Localisation and Industry Standard Association), Translation Memory Survey, 2004.

Maia, B. "Some languages are more equal than others. Training translators in terminology and information retrieval using comparable and parallel corpora." *Corpora in Translator Education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St Jerome, 2003. 43-53.

MeLLANGE Corpora & E-learning Survey Results. ITI Bulletin. 2006.

Marco Borillo, J., van Lawick, H. "Using corpora and retrieval software as a source of materials for the translation classroom." *Corpus Use and Translating*. Ed. A. Beeby, P. Rodríguez-Inés and P. Sánchez-Gijón. Amsterdam/Philadelphia: John Benjamins, 2009. 9-28.

Olohan, M. *Introducing Corpora in Translation Studies*. Routledge, 2004.

Optimale. Survey of translation technologies and tools teaching, 2012.



PACTE Group. "Building a translation competence model." *Triangulating translation: Perspectives in process oriented research*. Ed. F. Alves. Amsterdam: John Benjamins, 2003. 43-66.

Pearson, J. "Electronic texts and concordances in the translation classroom." *Teanga* 16 (1996): 86-96.

Pearson, J. "Teaching terminology using electronic resources." *Multilingual Corpora in Teaching and Research*. Ed. S.P. Botley, A.M. McEnery, and A. Wilson, 2000. 92-105.

Pearson, J. "Using Parallel Texts in Translation Training Environment." *Corpora in Translator education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St. Jerome, 2003. 15-24.

Peters, C., Picchi, E. "Bilingual reference corpora for translators and translation studies." *Unity in Diversity? Current Trends in Translation Studies*. Ed: L. Bowker, M. Cronin, D. Kenny and J. Pearson. Manchester: St. Jerome, 1998. 91-100.

Picton, A., Fontanet, M., Pulitano, D., Maradan, M. "Corpora in Translation: addressing the Gap between the Scholars' and the Translators' Point of View." *CULT Conference, 26-29 May*, Alicante, 2015.

Rodríguez-Inès, P. "Evaluating the process and not just the product when using corpora in translator education." *Corpus Use and Translating*. Ed. A. Beeby, P. Rodríguez-Inés and P. Sánchez-Gijón. Amsterdam: John Benjamins, 2009. 129-149.

Rossi, C., Frérot, C., Falaise, A. "Integrating controlled corpus data in the classroom: a case-study of English NPs for French students in specialised translation." *Corpus-based studies on language varieties*, Linguistic insights, Bern: Peter Lang, 2016.

Sauron, V. "Les nouvelles technologies dans l'enseignement de la traduction : l'exemple de la traduction juridique." *Traduction spécialisée : pratiques, théories, formations*. Ed. E. Lavault. Bern: Peter Lang, 2007. 207-224.

Torres Domínguez, R. 2012 Use of Translation Technologies Survey, 2012.

TradOnline. Translation Business and Translation Survey, 2010/2011.

Varantola, K. "Translators and disposable corpora." *Corpora in Translator Education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St Jerome, 2003. 55-70.

Wilkinson, M. "Using a Specialized Corpus to Improve Translation Quality." *Translation Journal* 9.3 (2005).

Williams, I.A. "A translator's reference needs: dictionaries or parallel texts." *Target* 8 (1996): 277-299.

Zanettin, F. "Bilingual Comparable Corpora and the Training of Translators." *Meta* 43.4 (1998): 613-630.

Zanettin, F. "Swimming in words: corpora, translation, and language learning." *Learning with Corpora*. Ed. G. Aston. Athelstan, 2001. 177-197.

Zanettin, F. "Corpora in Translation Practice." *Proceedings of the LREC Workshop, Language Resources for Translation Work and Research*. 2002. 10-14.

Zaretskaya, A. Corpas Pastor, A. Seghiri, M. "Translators' requirements for translation technologies: a user survey." *Proceedings of the AIET17 Conference New Horizons in Translation and Interpreting Studies*, 2015.

Recebido em 04 janeiro de 2016

Aceito em 22 de fevereiro de 2016

Publicado em abril de 2016