Integrating corpus-based tools into translators' work environments: cognitive and professional implications

Sandrine Peraldi
sandrine.peraldi@ucd.ie
University College Dublin

Abstract: *Despite the growing recognition of corpora-induced benefits for translation quality, the use of corpus technology remains an exception in a professional setting. This article describes the results of a pilot study aimed at integrating corpus-based tools into the work environment of legal/financial translators. In particular, the paper examines resistance and success factors with a view to encouraging the progressive integration of corpus technology into translators' toolboxes.*

Keywords: *corpus-based translation, legal translation, comparable corpus, cognitive efforts, technology awareness.*

I. Introduction and problem statement

The translation industry has been undergoing a profound technological transformation for the past 30 years, although not without some difficulty. The integration (and sometimes imposition) of computer-assisted translation (CAT) and machine translation (MT) functionalities as translation working aids - which are nowadays widely deployed and generally well accepted among the industry - have been regarded for a long time as highly disruptive due to the increased cognitive efforts and drastic changes work practices that they require from translators.

In line with this, corpus-based tools and methodologies are nowadays also being increasingly incorporated into translators' training programmes as an effective means of developing several key competencies among future translators. Indeed, a plethora of researchers have demonstrated the relevance of corpora and especially that of parallel corpora in training highly qualified translators (Bowker 1998, Kübler 2014, Loock, 2016). However, despite this wide academic recognition of corpora-induced improvements to translation quality, the use of corpus-based tools alongside CAT and MT tools remains an exception in professional contexts (Picton & al. 2015, Frérot, 2016).

This article is therefore aimed at understanding the reasons behind the persistent invisibility of corpora in the industry while discussing the many implications of introducing corpus technology among professional translators. After reviewing the transformative implications of Corpus Linguistics (CL) especially in Translation Studies (TS) and, paradoxically, its lack of recognition in a professional setting, the author will present the methodology and the findings of a field experiment launched two years ago at the Translation Centre of the French Ministry of Finance (Remfort & Peraldi 2017). This pilot study [1]was designed to determine if these new technologies can be integrated into a genuine work environment without disrupting translators' work habits and increasing their cognitive efforts. The article will conclude with a discussion of the preliminary findings, in particular in relation to the of potential adoption of early training and data collection solutions to resolve potential resistance factors from professional translators.

---

1 The relatively low number of participants taking part in the experiment requires the author to remain cautious about the results of the study.

II. Object of study

2.1 Defining Corpus Linguistics

The rise of corpus-based methodologies and the concomitant development of sophisticated computational tools are considered by many researchers as a transformative turning point in linguistics from both a methodological and a theoretical perspective. Indeed, thanks to the availability of large collections of authentic texts gathered in an electronic form (i.e. discourse used in genuine communicative events), linguists are now able to gain access to observable and verifiable naturally occurring data. Researchers can now study *'examples of what people have actually said, rather than hypothesising about what they might or should say'* (Pearson & Bowker, 2002: 9),

Of course, relying on large collections of texts especially in relation to language description is not a new activity. But the advent of computers in the 1960s combined with the increased availability of machine-readable data enabled linguists to investigate much larger and more representative samples of language and to discover linguistic information they might not have noticed through intuition alone. Corpora act as an objective frame of reference providing lexical, semantic, syntactic and statistical evidence of language use.

This new empirical method can be applied to a wide range of disciplines. Examples include foreign language teaching (for example the introduction of new corpus-informed teaching methods/materials), terminological practices (with the semi-automated creation and input of multilingual databases or the development of technical writing assistance tools) and, generally speaking, any discipline aimed at or concerned with examining language use.

The combination of natural language processing and Corpus Linguistics (CL) techniques such as part-of-speech annotation, semantic tagging and named-entity recognition have even allowed researchers to move beyond the frontiers of linguistics into many other humanities such as digital history, geography and law. A good illustration of this is the increased development of NLP-based information extraction tools which, for example, help identify and extract key text features and structured information (such as titles, section headings, dates or even names of companies) in legal and regulatory texts (Bommarito & al., 2018:1). By improving the efficiency and accuracy of document analysis, these new tools help legal professionals make more informed decisions. Another example is the combined use of data analytics and digital arts to enable new readings and a deeper understanding of historical and social events. In the Industrial Memories project developed by Keane, Pine & Leavy (2017), data analytics techniques were

applied to the Ryan Report, a 2009 report commissioned by the Irish government to enquire into the extent and effects of child abuse in Irish institutions for children. The mapping and structuring of the information enabled the creation of digital representations of the people who knew about these practices and the displaying of '*hidden patterns in the Report, illustrating the system of abuse in action*'.

Whether these approaches involve sophisticated data-mining methodologies based on computational power or simpler corpus-based discourse analysis, they all rely on the semi-automated extraction of new patterns of knowledge with, as demonstrated above, strong impacts on many spheres of society and professional life.

2.2 A twofold impact on translation theory and teaching

More recently, Corpus Linguistics has also disrupted the emerging field of Translation Studies, prompting in particular researchers & academics to readjust their way of envisioning the discipline as well as their teaching practices, especially in a vocational environment. The first impact of CL is indeed of a theoretical and epistemological nature as exemplified by a quite abundant literature. For a very long time, translation research was traditionally source-oriented and therefore rather prescriptive: the dominant question among theorists revolved around the idealised and formalistic concepts of fidelity and equivalence and the quixotic search for the perfect translation. Because of its many commonalities with and sometimes dependencies on cultural studies, literary studies and particularly contrastive linguistics, the discipline has suffered from a lack of scientific recognition (Ramón García 2002: 395). It was not until the 1980s that the first descriptive, empirical and, most importantly, target-oriented branch of translation studies emerged to strongly anchor the discipline from a theoretical perspective. Indeed, the Manipulation School led by theorists such as Holmes (1988) and Toury (1980, 1995) advocated the need to approach translation from a cultural, historical and, most importantly, functional perspective in order to understand the complex nature of the translation process. The product-oriented nature of this new approach not only meant studying actual translations, but also the context in which these translations were produced, their value and impact on the targeted readership, and the cognitive processes of translators (Rosa 2016: 96).

As explained by Granger (2003: 18), this slow change of perspective in TS research compelled researchers to rely on textual models and large bodies of translated texts to single out the underlying features of the translation process. And this, of course, is where corpora came into play. By examining differences and commonalities between comparable corpora of translated and non-translated

texts, and especially recurring patterns and styles of translations among translators, theorists such as Baker (1993) demonstrated that translations are not epigonic versions of the source text but *'form a distinctive textual system within any target culture'* (2004: 28). Because target texts are constrained by another text initially articulated in a completely different 'languaculture' and because translators are usually influenced by the social status of the text and its readership, they tend to conform to or even exaggerate the stylistic, idiomatic and syntactic features of the target language. The unveiling of these translation universals through the study of large pieces of text enabled a move away from purely stylistic and semantic considerations towards an approach that viewed translation as a genuine and atypical communicative event and as a social process deserving of analysis in its own right (Breedveld, 2002: 9).

Because the borders between research and teaching/training are, of course, undoubtedly porous, the second impact of CL was of an educational nature. As mentioned previously, corpus enthusiasts started demonstrating the relevance of corpora in innovating teaching methods and especially in the competence building of future translators (Bowker 1998, Kübler 2014, Loock 2016). For example, the use of comparable corpora enables translation trainers to highlight differences in use and style between the source and the target language, thus allowing students to capture translation as a complex reading and writing process and to progressively identify and apply translation standards as well as global and local strategies. As demonstrated by Bowker (1998), corpora also help students find appropriate equivalents and collocates by providing a wider context to the translator, especially compared to more traditional methods such bilingual terminological resources. Enhanced creativity, better-informed translation strategies, adequate use of technical terminology, increased use of idiomatic phrases and an improved understanding of the source text are just some examples of the many benefits of turning to corpus linguistics in the classroom.

III. The invisibility of corpora among translators: a literature review

Despite the many advantages of using corpora in translation teaching, the use of corpus-based tools alongside CAT and MT tools remains, to our surprise, an exception in the professional context, demonstrating a considerable gap between the academic world and the industry. This is all the more surprising given that the use of corpora has also widely - although indirectly - impacted the profession, due

to the widespread use of translation memories and machine translation which are after all both based on the use of parallel corpora and electronic dictionaries.

While there is an abundant literature addressing the use of corpus technology in academia, the lack of a corresponding literature in professional contexts is already a strong indicator of the invisibility of corpora among translators. Bowker & Pastor (2015), Frankenberg-Garcia (2015) and more recently Frérot (2016) are some of the very few authors offering a panorama of corpus technology in the industry or, to be precise, the lack thereof. Among the few reasons advanced by the above researchers to explain this situation, we note in particular three 'invisibility' markers.

First, the uneven and disparate teaching of corpus-based TS in master's degrees (even among EMT[2] universities). Competence-based teaching is today a centrepiece of most education programmes as well as a strong determinant in tool dissemination and adoption. As an example, the latest EMT Competence Framework published in 2017 puts a very strong emphasis on the use of CAT and terminology-based tools, MT and post-editing devices, but only mentions once the concept of 'corpus-based tools. As stated by Boulton (2007), implementing any kind of data-driven approach often implies questioning and reimagining traditional roles between educators and students which might account for some of the resistance encountered among lecturers. But most importantly, today's Master programmes are first and foremost a reflection of current market trends and needs. The same observation has been echoed in lifelong learning and vocational training, although some recent and welcome initiatives stemming from the Directorate General for Translation (DGT) (which has started to organise in-house corpus-based training sessions) and the recent integration of a fully-fledged module in the KU Leuven translation summer school represent positive developments.

As a matter of fact, and this is our second invisibility marker, there is very little demand emanating from translation providers as regards the mastery of corpus technology, especially compared to computer-assisted and machine translation tools. According to the latest figures from the 2018 study of the European language industry conducted jointly by leading associations of professional translators (the EUATC, Elia, FIT Europe, GALA and LINDWeb), more than 50% of professional translators now resort to MT and more than 80% have integrated CAT tools in their work environment, thus clearly demonstrating the discipline has shifted from 'a predominantly humanist profession to an increasingly-technology

2 EMT stands for European Master's in Translation.
3 See Skill 16 in the Technology skills subset (<https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf>).

driven practice' (Koskinen & Ruokonen 2017: 8). The report, which offers a detailed overview of professional practices, prices pressure, market expectations, etc., does not mention at all the use of corpus tools, an observation already made by Bowker & Pastor and Frankenberg-Garcia four years ago in similar surveys. Whereas there always has been very strong market pressure as regards CAT tools, which has directly impacted the skills development and training of professional translators, there seems to be none for corpus-based tools. The reasons behind this lack of pressure are, however, not addressed. One wonders if this could be linked to a technological cognitive overload among translators or perhaps the ultra-domination of computer-assisted software which leaves no space for new tools?

Thirdly, and most importantly, there seems to be a widespread lack of knowledge about the very concept of corpus and about the full capabilities of corpus technology (Carratalá-Puertas 2015; Frérot, 2016). While professional translators have slowly recognised the usefulness of large collections of texts for terminographic purposes (through automated term extraction provided by CAT tools), they appear to know very little, if nothing at all, about the existence of specific tools such as concordancers or syntactic analysers and how these tools could offer more powerful solutions to translators. The author's very own experience while working on TransCert, an EU-funded project aimed at designing a European-wide certification scheme in translation (Budin & al., 2013; Peraldi, 2014), also supports this statement. As part of the project, a pool of 50 professional translators and university-based trainers was asked to assess the e-learning training programme designed by the Consortium. The introductory module on corpus technology was the only one categorised as a new feature compared to the other more classic and already mastered computer-assisted skills.

A fourth reason which has not yet been investigated in corpus-based translation activities might be the strong resistance to technology that has long characterised the translating community. As we know, the arrival on the market of the first translation memories and rule-based automated systems in the 1990s triggered heated debates for many years within the community on whether computer-based technologies should be considered as a threat or conversely as an efficient aid that should be added to translators' toolboxes. For example, Olohan (2011) investigated the interplay between translators and computer-assisted tools by analysing discussion threads on professional translators' forums (the topic being discussed concerned the recent launch of a new version of SDL Trados). Two specific social trends seemed to emerge from the textual analysis of

'resisting' translators' forum posts: that of 'technological determinism' and 'social determinism'. According to Olohan, some of the respondents felt compelled to use technology for the sake of anticipating *changes in the practices of translation agencies*' (2011: 354) although they were not convinced by its usefulness (which also confirms the market pressure mentioned earlier). Another group appears to believe these changes serve the interest of software and/or translation service providers rather than that of translators. The terms 'technological determinism' and 'social determinism' used by Olohan have typically a negative connotation and therefore seem to convey a particularly pessimistic view of translators' willingness or ability to evolve and adopt new practices. In reality, this needs to be balanced by the numerous efforts seen over the years by professional translators helping to accelerate the technological race undertaken by software editors and research labs. However, what is striking in this account is the strong ethical stance taken by the respondents who seem to think that technology is systematically being forced on them.

Nearly ten years later, the considerable progress made by statistical, hybrid and now neural MT engines, the smooth integration of MT functionalities in pre-existing CAT tools, and the growing competition that characterises the market together have progressively outweighed translators' reluctance to use these technologies. Yet, according to Cadwell, O'Brien & Teixeira (2018: 301), a significant segment of the community still displays some sort of technology averseness. In a recent study carried out in 2016, they sought to analyse the key factors leading to the adoption or non-adoption of machine translation among professionals. To do so, they surveyed 17 focus groups of Luxembourg-based DGT translators and 4 focus groups of 20 translators working for a UK-based translation service provider. All translators were asked to specify the frequency of their use of MT software and explain the main reasons leading them to adopt or reject automated translation. Very interestingly, the study demonstrated an important gap between DGT-based and non-institutional translators, with the latter being overall more reluctant to adopt MT and post-editing activities. As an illustration, here are some of the most popular reasons given by pro-MT DGT staff: "Because of a personal interest in technology", "Because the translator wants to contribute to the improvement of the MT system", "Because of MT's positive influence on a translator's abilities" (2018: 310).

These findings seem to support the idea that a translator's internal environment (usually the commissioning institution) plays a decisive role in terms of technology adoption. It also appears that DGT translators' overall more positive attitude towards technology is linked to its early-stage integration

in the text production workflow of EU legislation. Facilitated feedback and interactions with EU lawmakers or in-house engineers in charge of MT@EC (the Moses-based MT engine) seem to trigger a higher sense of empowerment and usefulness among employees by allowing them a direct impact on the quality of the legislative texts. In contrast, the UK-based respondents seem to display a more fatalistic and detached attitude towards MT adoption ("Because greater MT adoption is inevitable", "Because translators are required to use MT") which seem to support Olohan's claims in terms of human/CAT tool interplay.

As regards the reasons for not adopting machine translation regardless of translators' professional affiliations, more classic concerns are put forward such as time, efficiency and especially quality loss (due to poor raw output and terminological inconsistencies), a dislike of post-editing activities which are deemed off-putting, a general sense of devalued work and more specifically, on non-institutional translators' part, a general distrust of machines.

As interesting as these results are, one wonders to what extent these resistance factors can be generalised to the lack of adoption of corpus technology. Indeed, machine translation and the inherent post-editing activities are considered a highly disruptive activity for translators. The partial automation of the translation process combined with enforced segmentation which characterises CAT and MT tools strongly impacts the translation strategies implemented by translators. Previous studies (Christensen & Schjoldager, 2010; O'Brien, 2012; Martikainen & Kübler, 2016) combining think-aloud protocols, quality assessment, keyboard logging and eye tracking showed that MT disturbs the translators reading and writing process (e.g. tendency to focus on individual segments at the expense of text coherence, to trust the machine instead of making microstrategic decisions in terms of style, lexis and readability, etc.). But most importantly, automated translation triggers extremely important emotional responses from translators as reflected in the answers described by Cadwell & *alii* (2018). The counter-intuitive nature of post-editing activities combined with the persistent and somewhat irrational belief that automated technologies might eventually replace human translators or at best considerably devalue their financial worth can be seen as key resisting factors in the context of MT. Interestingly, similar reactions are currently being displayed in the field of artificial intelligence. According to Juma (2016), society systematically tends to "reject new technologies when they substitute for, rather than augment, our humanity".

Conversely, corpus technology does not contribute to replacing human competence. Rather, it builds on human intuition by acting very often as a validation tool. Data-driven investigations enable researchers to verify

descriptions or hypotheses made first through introspection by allowing them to investigate genuine texts and to extract, for example, verified terminology, recurrent patterns or, on the contrary, atypical linguistic phenomena. The ability to rely on sound linguistic, semantic and technical data, combined with the efficiency and the speed with which requests can be performed, should in fact increase users' level of confidence in making informed decisions.

Furthermore, corpus-based activities seem to be consistent with translators' favourite tasks during the translation process. Koskinen & Ruokonen (2017) recently asked more than one hundred participants (either active professional translators or future graduates in TS) to write a short "love letter/break-up letter' to a tool, application or aspect of work of their choice". The purpose of this experiment, derived from usability studies, was to investigate translators' emotional narratives as a key factor in assessing their level of technology acceptance. The higher number of love letters addressed to translation technologies not only directly contradicted translators' supposed resistance to technology, but also demonstrated their particular fondness for search tools:

> […] there were seven love letters to traditional printed dictionaries, and four to research as such and the joy of discovering accurate equivalents or useful parallel texts. As described by one of the respondents, 'searching books and the internet for information is the best part of translation. (Koskinen & Ruokonen, 2017: 14)

Without wishing to generalise these findings to the entire translating community, this clearly shows that there is fertile ground for the use of corpus-based tools. It appears that translators' non-adoption of corpus technology is due more to a lack of corpus awareness and proficiency rather than an aversion to the technology. For example, many translators do not even realise that classic and daily-used functionalities such as concordance searches in translation memories, the use of parallel-corpus based applications such as Linguee, or even plain collocates search requests on web engines already fall into corpus-based proficiency (Picton & *alii* 2015) and could therefore be boosted by the use of more powerful tools. Interestingly, Dillon & Fraser (2006) had already reached the same conclusions when investigating translators' resistance to CAT tools.

Taken together, all these elements combined with the lack of research experiments aimed at analysing the potential benefits and disadvantages of using corpora specifically in a professional setting clearly call for greater collaboration between academia and professional representatives. This situation led to the implementation of the field experiment and pilot study described in the following sections.

IV. Methods

4.1 Project background

Bolstered by several years of successful collaboration, the author and the Translation Centre of the French Ministry of the Economy and Finance (MINEFI - Ministère des Économies et des Finances), spearheaded in particular by Julie Remfort (former deputy head of the centre), decided to conduct a one year-long experiment [3]to explore the possibility of integrating a corpus-based tool in a professional setting. This research project was aimed at investigating the 'usefulness' of corpus technology in terms of translation quality with a view to convincing in-house translators of its added value and encouraging its progressive integration into their work environment.

To put this project in context, the Ministry of Finance is one of the most important portfolios of the French government. It oversees the development, regulation and control of the economy (including industry, tourism, etc.), the preparation of the finance law (budget) and the drafting of taxation laws and employment policies, to name a few prerogatives. This means its translation services are asked to:

- perform translations in a very wide range of technical and highly specialised fields (ranging across the economy, finance, law and consumer regulations), with complex terminology and phraseology.
- produce high stake and high-quality translations as all documents officially 'bear the seal' of the Ministry,
- respect extremely tight deadlines by employing highly and multi-skilled translators for all of the above activities.

At the time of the experiment, the Centre's translation activities were structured around two main entities: 1) linguistic services with 3 in-house proof-readers and 12 in-house translators, and 2) linguistic, logistic and computational support whose role is to ensure terminological consistency while managing translation memories, indeed the Centre 'juggles' between several CAT software packages. Lastly, as displayed in the organisational chart, a specific service is dedicated to populating linguistic resources so that in-house translators can focus exclusively on the translation process.

─────────

3 The pilot study was conducted in the context of ISIT's Applied Research Projects.
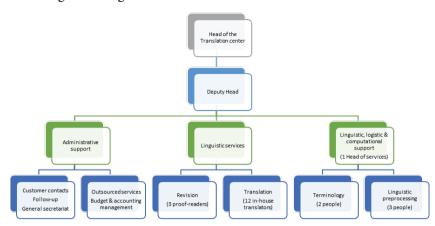
Figure 1: Organisation chart of MINEFI Translation Centre

Linguistic services are, however, often required to translate non-recurring and non-documented topics for which the use of translation memories is of no use. Surprisingly, there was no in-house jurist to clarify problematic concepts or provide insights based on comparative law. Thus, the adoption of new and innovative approaches, including the use of corpus technology, could be of utmost importance to help streamline the very strong activity of the centre as well as maintain translation quality.

4.2 Project design

Because of the strong applied and vocational nature of this study, it was decided to implement a field experiment that would reproduce as closely as possible the real-life working conditions of the Centre's translators. The experiment was therefore structured around the implementation of a two-hour long translation test carried out with the help of corpus technology and specific corpora in order to assess their usefulness.

More specifically, the project focused exclusively on assessing the relevance and the ease of use of comparable corpora which are rarely used in corpus-based translation activities (except for retrieving terminological information). As translation memories already act as dynamic parallel corpora, introducing the latter would do little to help translators in their daily activities. So far, comparable corpora have mainly been used to investigate the syntactic and lexical differences between translated and naturally occurring texts (Baker 2004, Zanettin 2013, Gallego-Hernandez 2016). The present research field is therefore legitimate and original in that it both attends to pragmatic needs and expands the current body of theoretical knowledge.

4.2.1 Corpus design

The first project milestone was the building of a comparable and multilingual corpus to be tested by the MINEFI in-house translators. The implementation of a real-life and company-based research project meant designing a relatively non-intrusive experience for the translators willing to take part in the experiment. As corpus design is a particularly time-consuming and demanding activity, it was decided to exempt the participants from building the corpus themselves to focus more specifically on the 'tool handling' phase. However, this raised questions and concerns inherent to data collection which will be discussed in section 5 of this article.

The selected field of application was trademark licence agreements. The subject, defined as a legal niche, gathers many of the recurring issues encountered by MINEFI translators. It displays a high level of specialisation and technicality and is a non-recurring topic.

The selected tool was Sketch Engine (SkE). The software developed by Kilgariff & *alii* (2014) is a concordancer with a built-in syntactic analyser that offered powerful text analysis functionalities. The relative user-friendliness of the interface, the availability of multiple ready-to-use corpora combined with the possibility of designing large tailored corpora directly influenced this choice. The tool is online which allows for easy deployment in any working and computational environment. However, this feature has also proven quite problematic as all the translated texts were highly sensitive documents and needed to be anonymised, which was particularly time-consuming and a clear drawback for the translation centre.

Compiling legal corpora in the context of legal translation is not a new activity. Prime examples of this approach are JuriGenT, a Dutch/Spanish legal database (Vanden Bulcke & De Groote, 2016), JudGentt, a translation-oriented glossary for criminal court translators (Borja-Albi & García-Izquierdo, 2016) and the EU-funded project QUALETRA (Kockaert, Peraldi 2014) aimed at providing a multilingual database in criminal proceedings. However, all these large-scale projects were mainly aimed at populating printed or electronic terminological knowledge bases, not using corpora as a translation aid.

Thus, three different multilingual sub-corpora (of an explanatory and phraseological nature) were compiled for this project. This threefold approach in particular assists in identifying the different stages of the translation process where corpus technology could prove to be most useful. As exemplified by Candel (2001) and Peraldi (2011, 2016), sub-corpora enable users to target in a timely manner specific and tailored information, although they multiplies

the number of resources that need to be managed by translators. All corpora were compiled according to the usual design criteria used in CL (Pearson & Bowker 2002, Peraldi 2016), such as representativeness, genre, corpus objectives, communicative settings, date of publication, source reliability, etc. although the size of the corpus proved to be by far the most problematic design criterion due to confidentiality and accessibility issues. Particular attention was also paid to prioritising integral and non-translated materials.

The first corpus, an English explanatory corpus, was designed as a documentary tool to help translators familiarise themselves with the subject field in the source language. It mainly comprises legal guidebooks, national legislation, official texts, summary files, etc.; in other words, documents belonging to an expert-to-expert or an expert-to-initiates communicative setting (Pearson, 1998) that allow for highly specialised definitions and technical explanations around major concepts. Despite the diversity of sources, the corpus remains surprisingly limited with 312,858 tokens. The main difficulty here resided in the very limited access to primary sources in UK law (due to its unwritten nature). Most documents emanated from the UK Intellectual Property Office.

The compiling of the French explanatory corpus followed a similar approach, but with a stronger focus on identifying potential equivalents. The very nature of French Law (which relies on a Civil Code and a Commercial code) greatly facilitated the corpus design as is evidenced by the number of collected tokens (608,332). The main difficulty here resided in understanding and representing the different legal hierarchies that govern French law. Indeed, according to Kelsen's Pure Theory of Law (1934), all legal systems are ruled by a hierarchy of norms, according to which specific legal texts prevail over others (for example, the French Constitution prevails over national law). One sees here the complexity of integrating relevant data into a specialised corpus as it has a direct impact on the representativeness of subject field and language variety.

Lastly, the French phraseological corpus, which barely amounts to 90,000 tokens, follows a distinct dynamic. It is aimed at helping translators identify and reproduce the typical collocates, idiomatic expressions and jurisprudential style of legal documents. The corpus is therefore exclusively comprised of trademark licence agreements, either in the form of templates or anonymised translations produced by the centre. In any field other than jurilinguistics, the smallness of the corpus would have been a major obstacle in terms of representativeness and linguistic validity. However, in the present case, the issue is counterbalanced by the very nature of the subject field. As stated by Bhatia, Langton & Lung (2004), legal discourse appears to be such a standardised genre (with particularly strong

formulaic, syntactic and stylistic constraints) that it is somehow characterised by an over-representation of typical linguistic phenomena (such as specific collocates, frozen expressions, jurisprudential style, etc.), thus allowing for the use of small corpora.

4.2.2 The experimental protocol

The literature review clearly demonstrated that introducing new technology always foreshadows significant changes in translators' work environments. For example, constantly switching from one tool to another entails an increase in the users' cognitive efforts. The Translation Centre was therefore particularly interested in determining in a very precise manner which specific features of corpus technology could address the very specific needs of professional translators (and especially those needs which are not currently addressed by the usual computer-assisted tools) while minimising as much as possible their level of disruption. The experiment was therefore aimed at answering three pragmatic questions:

- Which functionalities are most relevant during the translation process?
- At what stage of the translation process should corpus technology be used?
- What type of translation problems can be addressed?

An experimental protocol was designed to gather the views and reactions of in-house translators using Sketch Engine, while also assessing the quality of the target texts. The protocol was also built around a hypothetic-deductive approach as regards potential benefits of corpus technology. The use of corpus technology should i) enable a better understanding of the technical field involved, ii) significantly save time and improve translation quality, iii) allow targeted and efficient linguistic queries during the translation process, and iv) allow for fast and easy adoption. Although the underlying approach of this pilot study is pro-corpus, it was equally important to unveil sources of resistance to corpus technology to start reflecting on potential solutions for better integrating CL tools in a working environment.

The first stage of the protocol consisted of training the participants in mastering Sketch Engine. Four 2-hour hands-on sessions and the drafting and dissemination of a 15-page translator-oriented handbook was deemed sufficient to enable translators to adequately master the software. The handbook was, however, tailored to the specific needs of translators by using carefully drafted examples and exercises. Each training session was also followed by a group

discussion to start the process of compiling the participants' first impressions and concerns.

The second stage also adopted a combined approach involving questionnaires, semi-directed interviews and again group discussions. The questionnaire was aimed at identifying specific benefits of using corpus technology (such as finding collocates, equivalents, etc.) at precise moments of the translation process (pre-translation documentary phase, proofreading, etc.) Participants were asked to rate (on a scale of five) the usefulness of Sketch Engine for each specific task and translation phase and to describe its strengths and weaknesses through the use of open questions to allow for more personal and specific comments. Semi-directed interviews were used as follow-ups in order to delve further into the participants' written answers.

Lastly, because of the strong technicality of the subject, the quality assessment of the tests was assigned to a jurilinguist working in the Ministry, therefore following a purely holistic assessment approach (Gardy 2016). Specific focus was drawn on text readability and coherence, the use of appropriate terminology and idiomaticity.

At this stage, it should be highlighted that 4 translators and 1 translator/terminologist, out of the 17 full-time members of the translating team, agreed to take part in the experiment. The workload of the translation services is such that the Centre's management deemed it very difficult to ask all staff members to engage in the full experiment (induction sessions, training in Sketch Engine, test, survey and discussions). As already mentioned, although the field experiment had been designed from the very beginning as a pilot project, the importance of interpreting the present results with great caution is clearly acknowledged. All participants were experienced MINEFI translators and considered themselves to be non-specialists in the field of trademark agreements.

V. Results and discussion

5.1 Identifying key functionalities.
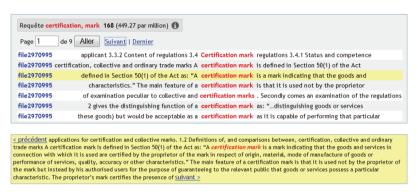
The analysis of the surveys combined with individual interviews enabled four particularly beneficial functionalities used at specific stages of the translation process to be identified. These functionalities will, however, be analysed in the light of the comments and criticisms made by participants and the external jurilinguist in order to identify potential resistance factors.

5.1.1 Understanding key concepts in the source language

As anticipated, terminological research in both languages and more particularly the use of well-known functionalities such as *word list* (extraction of the most recurrent terminological items) or the analysis of specific concordance lines (context search) were particularly praised by the participants, with scores systematically ranging between 4 and 5.

The identification of definitional information thanks to the use of linguistic or textual markers was also highlighted as a key feature. Linguistic markers are "observable text features identified through corpus analysis, signalling the kind of relations between lexical items used in building terminologies […] or relations between text segments involved in discourse coherence" (Condamines & Péry-Woodley, 2007:2) In the following screenshot, the definitional marker 'is a/ is a sort of/is a kind of' is used to gather information around the concept of *certification mark*.

Requête **certification, mark** 168 (449.27 par million) ⓘ

Page 1 de 9 [Aller] [Suivant] | [Dernier]

file2970995	applicant 3.3.2 Content of regulations 3.4 **Certification mark** regulations 3.4.1 Status and competence
file2970995	certification, collective and ordinary trade marks A **certification mark** is defined in Section 50(1) of the Act
file2970995	defined in Section 50(1) of the Act as: "A **certification mark** is a mark indicating that the goods and
file2970995	characteristics." The main feature of a **certification mark** is that it is used not by the proprietor
file2970995	of examination peculiar to collective and **certification marks** . Secondly comes an examination of the regulations
file2970995	2 gives the distinguishing function of a **certification mark** as: "...distinguishing goods or services
file2970995	these goods) but would be acceptable as a **certification mark** as it is capable of performing that particular

< précédent applications for certification and collective marks. 1.2 Definitions of, and comparisons between, certification, collective and ordinary trade marks A certification mark is defined in Section 50(1) of the Act as: "A *certification mark* is a mark indicating that the goods and services in connection with which it is used are certified by the proprietor of the mark in respect of origin, material, mode of manufacture of goods or performance of services, quality, accuracy or other characteristics." The main feature of a certification mark is that it is used not by the proprietor of the mark but instead by his authorised users for the purpose of guaranteeing to the relevant public that goods or services possess a particular characteristic. The proprietor's mark certifies the presence of suivant >

A quick analysis of the context enabled the participants to retrieve not only a full definition of the concept (*A certification mark is a mark indicating that the goods…*) but also reliable and contextualised information and avoided the need for them to browse through multiple terminological resources.

It is worth noting that these document-handling skills are a common competence that translators typically master to a high level. Participants therefore seem to award high scores to functionalities belonging to their 'comfort zone'. As depicted by Picton & al. (2015), translators naturally turn to search engines and use Google as a 'mega corpus' for encyclopaedic information, domain familiarisation, etc. They are nonetheless not used to using reliable and tailored corpora instead (bringing up again the issue of corpus design). Therefore, the central question that seems to emerge here is how to initiate a shift in translators' search habits, rather than making them acquire a new set of skills.

5.1.2 Exploring phraseology

The search for idiomaticity through the exploration of phraseology is one of the core functionalities provided by any concordancer. One of Sketch Engine's strengths is to provide a syntactically categorised display of the recurrent collocates surrounding a specific term. In the screenshot below, all collocates of the French term *contrat* (contract) are classified according to their grammatical function in a single window, thus allowing the user to grasp at once the most typical jurisprudential expressions.

contrat *(noun)* Corpus phraséologique FR fréquence = 1,074 (9,766.03 par million)

modifier			sujet_de			et_ou			pp_de		
	16	0.01		376	0.35		106	0.10		105	0.10
opposable	3	12.26	être +	135	12.10	tort	8	11.15	licence	76	13.02
valide	2	11.67	résilier	18	10.50	résiliation	10	11.14	*contrat de licence*		
"	4	10.98	conclure	17	10.38	*terme du présent contrat ou de la résiliation de la licence*			sous-licence	17	12.04
autre	2	10.95	*présent contrat est conclu*			partie	10	11.01	*contrats de sous-licence ..*		
			prendre	16	10.28	terme	8	10.84	sous-traitance	2	9.26
			Contrat prend			concessionnaire	6	10.68	mariage	2	9.26
			pouvoir	18	10.15	redevance	4	10.11	achat	2	9.17
			Contrat peut être			concédant	4	9.92			
			régir	12	9.97	an	3	9.81	**objet_de**		
			avoir	10	9.39	moi|mois	3	9.78		99	0.09
			soumettre	8	9.36	destruction	2	9.24	résilier	22	12.49
			emporter	6	9.00	concession	2	9.21	*peut résilier le contrat*		
			savoir	6	8.97	changement	2	9.20	present	15	12.07
			conférer	5	8.73	annexe	2	9.20	*. Le present contrat*		
			si	5	8.73	autorisation	2	9.16	conclure	16	11.95
			constituer	5	8.71	>	2	9.16	signer	12	11.74
			résoudre	4	8.43	tiers	2	9.12	rendre	6	10.68
			annuler	4	8.42	renouvellement	2	9.10	durer	4	10.24
			établir	4	8.42	usage	2	8.98	céder	3	9.81

Although both the author and the jurilinguist noted a significant rise in terms of text fluency and idiomaticity compared to previous translations done by the same team of translators, participants' impressions were paradoxically more mixed. One participant acknowledged that the concordancer enabled him/her to use combinations that he/she might not have thought of on his/her own and offered richer and more diverse resources. But three participants gave a score of only 3, explaining that they mainly used Word Sketch to confirm their intuition rather than explore new phraseology and come up with more original resources. Interestingly, one of these three participants suggested using the tool at a later stage of the translation process, during the proofreading phase, to polish the final version of the translation.
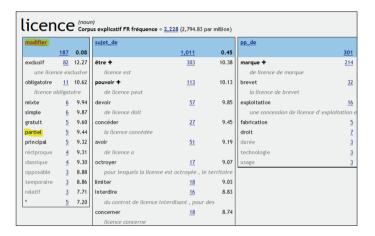
5.1.3 Finding equivalents

The use of corpus technology also enabled in-house translators to find specific equivalents that could neither be found in traditional terminological resources (in this instance, in MINEFI terminological database), nor on the internet in reliable documents. The interesting thing is that translation solutions were found

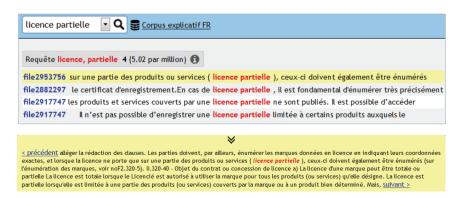by drawing up specific hypotheses and then confirming them by exploring the corpus.

For example, the concept 'limited licence' appears not to be associated with a registered equivalent. One of the participants assumed that the term could be translated by one of these three possibilities:

- *licence limitée* (back translation: limited licence)
- *licence non exclusive* (non exclusive licence)
- *licence restreinte* (restrained translation).

The approach consisted of researching modifying collocates (in this case, adjectives) in the target language, again using Word Sketch:

The translator spotted instantly the expression 'licence partielle' (partial licence). A quick context search enabled him/her to find the concept definition and eventually validate the equivalent, by cross-referencing the term in the source and target language.

This functionality allows the user to considerably speed up the search for problematic or unknown equivalents. It also turns out to be intellectually very satisfying as it activates the typical investigation skills praised by translators, as already pointed out by Koskinen & Ruokonen (2017).

These two components both seem to explain the high scores (systematically 4 or 5) granted by all participants. Two of them, however, conceded that mastering the logic behind the use of Sketch Engine was not easy at first and entailed a cognitive shift in their way of thinking. Corpus linguistics is indeed based on inductive reasoning. One has to draw on careful data observation to reach a conclusion (in this case, a solution to a specific translation problem). This 'noticing' ability is not a typical competence acquired by translators who tend to either use their memory/experience, either rely on 'turnkey' solutions delivered by translation memories, machine translation or terminological databases.

5.1.4 Choosing between synonyms

The last functionality identified as particularly useful was Sketch Diff. Word Sketch Difference is initially used to '*compare and contrast two words by analysing their collocations and by displaying the collocates divided into categories based on grammatical relations*' (Sketch Engine website, last updated 2018). In the present experiment, the functionality was intentionally misused to help translators choose between several synonyms.

For example, the English term *trade name* can be, depending on the context, translated as *nom commercial* or *désignation commerciale*. Both terms were submitted in the software. Sketch Engine assigned the green colour to *nom* and the red colour to *désignation*. Green collocates are more closely related to the first term and red collocates to the second. The stronger the colour, the more usual it is to use a specific combination of words. As displayed in the screenshot, the corpus exclusively points towards the use of *nom commercial* in the context of trademark licence agreements.

As we can see, corpus technology again allows quick and informed decisions. As pointed out in more informal discussions, this efficiency feature was particularly appreciated by translators who systematically operate in a just-in-time working environment.

5.1.5 Summary

The experiment enabled us to identify four very specific functionalities that meet diverse needs (searching for definitions and validating collocates, equivalents and synonyms) at varied stages of the translation process (source text decoding, translation and, potentially, final proofreading). However, all these functionalities seem to share one strong commonality. The concordancer indeed appears to be particularly useful and appreciated by all participants when it acts as a validation tool for pre-existing translation hypotheses. Translators first rely on their intuition/experience and on more traditional resources (glossaries, translation memories, etc.) to come up with a potential solution and then turn to corpus technology in a subsequent phase to consolidate their choice. The fact that the translators were not in the habit of using corpus technology and its novelty aspect *versus* their years of training and professional experience with more traditional tools can easily account for their 'second-line' use of corpus tools. The apparent correlation between high scores given to more comfortable functionalities would also seem to parallel this finding.

Consequently, the question remains as to how corpus technology can help translators formulate new translation hypotheses and be used much sooner in the translation process and not just as a validation aid.

5.2 General impressions and limitations

Despite some positive aspects highlighted throughout the experiment, several other factors also gave cause for concern. One of the highest resistance factors appeared to be the issue of compiling reliable and efficient corpora. This concern appeared very early in the group discussions. It was categorically stated that corpus

design was incompatible with translators' heavy workload. The participants however adopted a more positive and relaxed attitude towards corpus technology as soon as they were presented with the possibility of using an embedded corpus-building tool. The WebBootCaT functionality for example automatically creates corpora using web pages, by allowing the user to specify seed words or specific URLs on a given topic. Although this was only presented as a quick backup solution (due to the lack of specific design criteria to ensure representativeness and reliable data), the time-saving arguments almost immediately regained their momentum during the training session.

Participants also bemoaned the lack of usability of the tool which appeared to be too complex and tailored-made for linguists and not translators (particularly the terminology used in the software such as n-grams, lemmas, etc.). The recent redesign and simplification of the interface will probably solve some of these concerns. However, designing and providing training in a translator-oriented toolkit proved to be essential to mastering the tool quickly.

Furthermore, the impossibility of displaying extended contexts and browsing through the entirety of the text was felt as a major drawback. Fortunately, tools other than Sketch Engine offer this possibility. This ergonomic constraint might also partly explain the very low scores given by the participants as regards domain familiarisation. Most participants did not feel they could particularly explore the subject field nor the typical phraseology to gain a better understanding of the subject despite that fact that this is a recurring benefit highlighted by many academics specialising in corpus-based translation studies. As with CAT tools, it seems that the fragmented information provided through the exploration of concordance lines and lists of collocates disturbs the reading process and the processing of specialised information by translators.

Finally, although all translations were deemed more fluent and idiomatic by the jurilinguist, there was no particular gain in terms of time. All translators admitted they had difficulties in changing their reflective mode and that they would need more time to fully adapt to corpus technology.

In the final question of the survey, respondents were asked if they would consider using corpus technology in their working environment. Two translators answered yes and three of them maybe. Despite the very reduced size of the focus group, these answers can be considered quite promising, provided that MINEFI in-house translators are given the opportunity to test the tool on a long-term basis and that strong incentives are offered by the different heads of units. Indeed, the voluntary nature of this field experiment and the extremely short testing period did not allow the translators to integrate the use of corpora into their working

environment, thus meaning that slow progress was made in terms of changing their working and cognitive habits.

These preliminary results clearly call for a much broader and longer experiment with extended focus groups in order for the findings to be generalised with certainty. A forthcoming experiment reproducing similar conditions to the ones described by Caldwell & al. (2018), entailing the testing of corpus technology over several months would allow the author to monitor and analyse with greater precision translators' progressive acquisition of corpus-based skills and their efforts to accommodate this technology. The present pilot study also established that translators tend to rely on already acquired competences and tools. The design of a long-term experiment involving, on one hand, young translation graduates (trained in corpus technology) and, on the other, experienced translators who have only recently become acquainted with corpus tools would also help determine if the early adoption and use of these skills could have a positive impact in terms of cognitive efforts and work practices.

Lastly, a much more consistent and also time-dependent quality assessment with clear and objective quality evaluation criteria is needed to unequivocally establish the benefits of corpus technology in terms of quality, time and possibly costs in order to raise the awareness among both professional translators and translation industry decision makers. The recent integration of a Sketch Engine plugin within Trados Studio, however, confirms the growing market interest in corpus tools.

VI. Conclusion: towards a redefinition of the profession

The primary challenge of legal translators is to communicate in a target 'languaculture' the subtleties and intricacies of a completely different legal/judicial system characterised by its very own and unique historical and social evolution. Just as the cultures, traditions and languages of countries have evolved differently and subsequently diverged from others, so too have their legal systems. Legal systems reflect the way in which a community of speakers of a certain language perceives the world and creates concepts in order to understand, categorise and name this reality, thus making transpositions from one system to another extremely complicated.

These intercultural differences give rise to very specific linguistic issues such as conceptual non-equivalence or indeterminacy, untranslatable matter, overlapping between European and national concepts, etc. Added to the technicality and the typical jurisprudential style that characterises legal texts, translating such texts is a true endeavour.

This research project clearly showed that the use of well-tailored and appropriate corpora can be particularly relevant to the field of legal translation. Indeed, it was demonstrated that, when incorporated at specific stages of the translation process, it could benefit in-house translators quite quickly in terms of terminological coherence, improved idiomaticity and in finding non-documented equivalents. The relatively quick mastering of Sketch Engine among participants combined with the relatively high scores given in the survey suggest that the resistance factors triggered by corpus technology are manageable, as established by the literature review, at least compared to the disruptive effects of machine translation. It also appears that the use of corpora can complement quite efficiently the deficiencies of more traditional resources (such as terminological databases and translation memories), especially by acting as a powerful validation tool in terms of terminology, phraseology and translation hypothesis.

However, despite these advantages, it also appears that corpus-based technology entails an important cognitive shift among professional translators, not only in terms of changing working habits and tool selection, but also with regard to their reflective mode. Despite the preliminary nature of the present findings, it seems that translators naturally turn to 'comfortable' solutions and tools as a first-line process and instead use Sketch Engine as a validation tool rather than a decision-making tool that could increase creativity and translation quality. Given the relative novelty and lack of awareness of corpus technology in the industry, it seems to the author that a systematic and early integration of corpus-oriented skills in curricula and vocational training is needed to foster among translators-to-be a much wider and effortless use of data-driven techniques in their working environment and habits. As much as market incentive is essential to tool adoption, new graduates trained in the latest tools and methodologies can also bring about slow but steady and significant changes in the industry.

Lastly, the dissemination of corpus technology can only be achieved in the foreseeable future if the issue of data accessibility and digitisation are tackled in an efficient way. Translators who are by nature overloaded with work therefore need to be presented with reliable and easily accessible resources if they are to invest time and effort in mastering new software. The structure of the MINEFI Translation Centre (where all the pre-processing of documents is carried out by a distinct unit) combined with the respondents' concerns in terms of corpus design clearly showed that the building of large and reliable corpora should ideally be undertaken by a dedicated person or team. Furthermore, corpus design is a particularly difficult task which entails both a strong knowledge of the data being retrieved and compiled and a deep understanding of the main difficulties encountered by translators if the aim is to offer them perfectly tailored resources.

At a time when the retrieval of terminological data is being increasingly automated[4] and translators increasingly rely on computer-aided technologies, the focus of terminological activities is not so much on compiling terminological records, but on building reliable sources to feed these different tools. The expertise, the attention to detail and demand for quality that characterises terminologists would be perfectly suited to designing quality corpora. We therefore call for the evolution of terminologists' activities to embrace these new technological changes in order to meet the growing linguistic challenges faced by professional translators.

Acknowledgements

References

Baker, M. (1993). "Corpus Linguistics and Translation Studies. Implications and Applications". In M. Baker, G. Francis, and E. Tognini-Bonelli (eds). *Text and Technology*. Amsterdam & Philadelphia: Benjamins, p. 233-250.

Baker, M. (2004). "The treatment of variation in corpus-based translation studies". *Language Matters*, 35(1): 28-38.

Bhatia, V. K.; Langton, N. & J. Lung (2004). "Legal discourse: Opportunities and threats for corpus linguistics". In U. Connor & T. A. Upton (Eds.). *Discourse in the professions: Perspectives from corpus linguistics*. Amsterdam: John Benjamins, p. 203-231.

Borja Albi, A. & I. García-Izquierdo (2016). "Web-based tools and resources for legal translators: the JudGENTT translation-oriented glossaries for criminal courts translators". *Onomázein*, 33: 226-250.

Boulton, A. (2007). "Esprit de corpus : Promouvoir l'exploitation de corpus en apprentissage des langues". *Texte et Corpus*, 3 : 37-46.

---

4 For example, the software provider Lingua et Machina can retrieve 400,000 legal terms and their corresponding semantic relations.

Bowker, L. (1998). "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study". *Meta*, 43 (4): 631-651.

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

Bowker, L. & J. Pearson (2002). "Working with Specialized Languages". *A Practical Guide to Using Corpora*. London, Routledge.

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

Bommarito, M. J.; Katz, D. M. & E. Detterman (2018). *LexNLP: Natural Language Processing and Information Extraction For Legal and Regulatory Texts*. Available at SSRN: <https://ssrn.com/abstract=3192101>.

Breedveld, H. (2002). "Writing and revising process in professional translation". *Across languages and cultures*, 3(1): 91-100.

Budin, G.; Krajcso, Z. & A. Lommel (2013). "The TransCert project: ensuring that transnational translator certification meets stakeholder needs". *The International Journal for Translation & Interpreting Research*, 5(1): 143-155.

Carratala-Puertas, L. (2015). "Corpus y traducción profesional, una relación tan omnipresente como invisible". CULT Conference, 26-29 May 2015, Alicante.

Cadwell, P.; O'Brien, S. & C. S.C. Teixeira (2018). "Resistance and accommodation: factors for the (non-) adoption of machine translation among professional translators". *Perspectives*, 26 (3): 301-321.

Candel, D.; Marchaudon, P. & V. Tolédano (2001). "Le groupe nominal dans l'activité définitoire des scientifiques". *Langue et discours spécialisés*, Revue française de linguistique appliquée, 6(2) : 17-28.

Christensen, T. P. & A. Schjoldager (2010). "Translation-Memory (TM) Research: What Do We Know and How Do We Know It?" *Hermes - Journal of Language and Communication Studies*, 44: 89-101.

Condamines, A. & M-P. Péry-Woodley (2007). "Linguistic Markers of Lexical and Textual Relations in Technical Documents". In Alamargot, D.; Terrier, P. & Cellier, J.-M. (Eds.), *Written documents in the workplace. Studies in Writing*. Amsterdam: Elsevier, 3-16.

Dillon, S. & J. Fraser (2006). "Translators and TM: An investigation of translators' perceptions of translation memory adoption". *Machine Translation*, 20 (2): 67-79.

Frankenberg-Garcia, A. (2015). "Training translators to use corpora hands-on: challenges and reactions by a group of 13 students at a UK university". *Corpora*, 10 (2): 351-380.

Gardy, P. (2016). "L'évaluation en didactique de la traduction : un état des lieux. Didactic assessment in translation: a comprehensive update." *The Journal of Specialised Translation*, 26: 20-49.

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

Holmes, J. (1988). "The Name and Nature of Translation Studies". In J. Holmes (Ed.) *Translated!: Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi, 67-80.

Juma, C. (2016). *Innovation and Its Enemies: Why People Resist New Technologies*. New York: Oxford University Press.

Kilgarriff, A.; Baisa, V.; Bušta, J.; Jakubíček, M.; Kovář, V.; Michelfeit, J.; Rychlý, P., & V. Suchomel (2014). "The Sketch Engine: ten years on". *Lexicography*, 1: 7-36.

Kockaert, H. & S. Peraldi (2014). "Qualetra: Implementation of Directive 2010/64/EU on the right to interpretation and translation in criminal proceedings". *Proceedings of the XX World Congress of the International Federation of Translators (FIT)*, Berlin, August 4-6, 2014.

Koskinen, K. & M. Ruokonen (2017). "Love Letters or Hate Mail? Translators' technology acceptance in the light of their emotional narratives". In Kenny, D. (ed). *Human Issues in Translation Technology*. London: Routledge

Kübler, N. (2014). *Traduction pragmatique, linguistique de corpus, traducteur : un ménage à trois explosif ?* Tralogy II Conference [Online], Session 3 - Machine and Human Translation: Finding the Fit? / TA et Biotraduction, updated: 02/06/2014

Loock, R. (2016). *La traductologie de corpus*. Villeneuve-d'Ascq : Presses Universitaires Du Septentrion, 261 p.

O'Brien, S. (2012) "Translation as human-computer interaction". *Translation Spaces*, 1 (1): 101-122.

Olohan, M. (2011). "Translators and translation technology: The dance of agency". *Translation Studies*, 4(3): 342-357.

Martikainen, H. et N. Kübler (2016). "Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions".

ILCEA [En ligne], 27 | 2016, mis en ligne le 08 novembre 2016, consulté le 20 octobre 2018. URL: <http://journals.openedition.org/ilcea/3863>.

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

Pearson, J. (1998). *Terms in Context*. Amsterdam: John Benjamins Publishing Company. (Studies in Corpus Linguistics; 1).

Peraldi, S. (2014). "The TransCert project: Trans-European Voluntary Certification for Translators (Panel)". *Proceedings of the XXth World Congress of the International Federation of Translators (FIT)*, Berlin, August 4-6, 2014.

Peraldi, S. (2016). "La chimie organique : une science en mouvement". In P. Phillips-Batoma & F. Zhang (eds). *Translation as Innovation: Bridging the Sciences and the Humanities*. Dalkey Archive Press, 149-167.

Pine, E.; Leavy, S. & M. Keane (2017). "Re-reading the Ryan Report: Witnessing via and Close and Distant Reading". *Éire-Ireland*, 52 (1 & 2): 198-215.

Ramon Garcia, N. (2002). "Contrastive linguistics and translation studies interconnected: The corpus-based approach". *Linguistica Antverpiensia*, 1: 393-406.

Remfort, J. & S. Peraldi (2017). "Intégration des outils de traitement de corpus en contexte professionnel : étude de cas dans la traduction juridique et financière". *Translating Europe Forum: "New skills, New trends, New markets"*, DGT, European Commission, Brussels, November 6-8, 2017.

Termmerman, R. (2000). *Towards New Ways of Terminology Description. The sociocognitive approach*. Amsterdam/Philadelphie: John Benjamins.

Toury, G. (1980). *In search of a theory of translation*. Tel Aviv: Porter Institute.

Toury, G. (1995). *Descriptive Translation Studies and beyond*. Amsterdam & Philadelphia: Benjamins.

Vanden Bulcke, P. & C. De Groote (2016). JURIGENT, a different bilingual database of legal Dutch/Spanish. *CLINA*, 2(2): 15-38.

Zanettin, F. (2013). "Corpus Methods for Descriptive Translation Studies". In *Procedia-Social and Behavioral Sciences 95*, 20-32.