

## Surveying the use of CAT tools, terminology management systems and corpora among professional translators: General state of the art and adoption of corpus support by translator profile<sup>1</sup>

Heidi Verplaetse & An Lambrechts

KU Leuven, Belgium

---

### Abstract

Several international surveys show that language professionals are well acquainted with computer-assisted translation (CAT) tools as well as with (integrated) terminology management systems. However, corpora can be valuable in the translation process as well. To complement the previous surveys, we conducted a survey among Belgian and Dutch translation professionals which assessed their use of CAT tools, terminology management systems and corpora.<sup>2</sup> We distinguished the translator profiles of in-house and freelance translators. On the basis of the literature review we hypothesized that corpora, CAT tools and translation management systems are used more frequently by in-house translators. But, whereas the latter two are indeed more popular among the in-house translators of our survey, corpora are much less frequently used by both translator profiles in surprisingly equal measure. Parallel corpora are used most by the respondents. However, the distinction between a translation memory (TM) and a corpus is not always clear to professional translators in our survey. In order to increase familiarity with corpora among translators we plead for a broad implementation of corpora into CAT tools.

### Keywords

Corpora, CAT tools, terminology management systems, translation profession, translator profile

---

<sup>1</sup> We would like to thank both anonymous reviewers for their very useful suggestions and comments to improve the current article.

<sup>2</sup> The research for this paper was conducted under project fund STG/15/040/IOF for the project entitled *Impact of the use of corpora and CAT tools on translation quality and efficiency for translation of specialized texts*, granted by KU Leuven Internal Funds.

## 1. Introduction

Survey research is a frequently used instrument to assess the use of corpora, CAT tools and terminology management systems, both in Europe and worldwide (Cf. section 2. Survey research: an overview). Our small-scale survey conducted among Belgian and Dutch professional translators confirms previous surveys, indicating a low level of familiarity with corpora (Zaretskaya, Corpas Pastor, & Seghiri, 2015, p. 250; Zaretskaya, Corpas Pastor, & Seghiri, 2018, p. 47). This can be attributed to the lack of training for the use of corpus tools (e.g., *WordSmith*, *AntConc*, *Sketch Engine*) among future translators, the lack of ready-made specialized corpora (Wilkinson, 2010), the lack of awareness of the potential benefits of corpora “as a supplement to other resources and references” among future translators (Frankenberg-Garcia, 2015, p. 354), or the lack of translator-oriented corpus-tools (cf. Fantinuoli, 2016). The compilation and exploitation of corpora is also time consuming and corpora do not provide productivity gains immediately (Bowker, 2004, p. 240). This leads to the assumption that, in order to promote the use of corpora in the translators’ workflow, it may be useful to have “an easy-to-use program for compiling and managing parallel and comparable corpora” (Zaretskaya *et al.*, 2015, p. 253).<sup>3</sup> In addition, translators do not necessarily need ‘clean’ and annotated corpora, contrary to linguists, who mainly want to draw scientific conclusions from the corpus data (Loock, 2016, p. 18). The time investment for corpus compilation and exploitation for academic purposes might be circumvented for professional translators’ use of corpora. In the future, the integration of corpora into CAT tools could address these drawbacks (Bernardini & Castagnoli, 2008, p. 52). Integrating corpora in a CAT tool will also sensitize translation professionals to the potential benefits of corpora for translation. These may include the addition of non-translated original native language data to the traditional translated TM-segments translators usually rely upon.

Contrary to corpora, TM technology is integrated as a central feature in CAT tools and the TM function is widely used by professional translators. But TM technology also entails disadvantages. Poor-quality TMs will negatively impact the global translation quality. Therefore, TM maintenance is of utmost importance.

For the purposes of the current paper, we want to look into the adoption of corpora by professional translators, as well as the adoption of the (more established) CAT tools and terminology management systems and establish any potential influence of different translator profiles.

The article is structured as follows. We start with an overview in section 2 of previous surveys conducted on a European and a global scale to assess the use of corpora, CAT tools and terminology management systems between 2004 and 2016. In reviewing the existing recent surveys, we paid special attention to the potential impact of translator profiles. This has prompted a hypothesis with a focus on in-house versus freelance translators and uptake of new technologies: we expected a higher frequency of corpus use among in-house translators compared to freelance translators among our respondents. As the proportions in the available surveys we review in section 2 are approximative indications for these profiles, we distinguished discrete categories for the in-house and freelance translators in our small-scale survey, related to different aspects of the use corpora, CAT tools and terminology management systems as translation aids.

---

<sup>3</sup> Depending on the scholar, a comparable corpus can either consist of (1) two single monolingual corpora, containing original texts (written in the language of the native speaker) in a particular language and translations in that language (Baker, 1995, p. 234) or (2) original texts in two or more languages (Johansson, 2007) or language varieties, cf. [http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus\\_ling/content/corpora/types/comparable.html](http://www.essex.ac.uk/linguistics/external/clmt/w3c/corpus_ling/content/corpora/types/comparable.html) matching in genre, moment of publication, etc. (Johansson, 2007, p. 9).

A second hypothesis concerns the type of corpus use. It may be expected that parallel corpora are far more popular among translators than other types of corpora. But comparative and monolingual corpora may impact translation quality positively as well. Therefore, we have also analyzed our survey data on reported corpus types.

A third hypothesis concerns the availability of corpora dependent on the translated language pairs and corpus tradition in those languages. We hypothesized that translators working with English will consult different types of corpora (incl. monolingual) more frequently in their translation practice than other translators.

The results of our survey data for Belgian and Dutch professional translators on the above hypotheses will be discussed in section 3. Section 3 starts with a comparison of some overlapping survey questions in another recent Belgian survey, which is mainly market-oriented and which was conducted shortly after we launched our survey. In section 3.3 we add a comparison of freelance and in-house translators' work practice concerning corpora, CAT tools and terminology management systems on the basis of the discrete categories for these respondent profiles. In section 3.4 we provide more fine-grained analyses of respondent data for corpus use by language combination and types of corpora on the basis of our small-scale survey.

Finally, we look at current and future possibilities of integrating corpora into CAT tools in section 4.

## **2. Survey research: an overview**

### **2.1. Surveying the use of corpora**

#### **2.1.1. Introduction and rationale**

We begin our overview of survey research into the use of electronic translation aids with the least obvious aid from professional translators' perspective, but the type we wish to focus on with this paper, as we believe it may fill an opportunity for better translation product quality, viz., that of corpus use. We focus on information regarding respondent occupation from the surveys we will review, as we hypothesize that the use of corpora as a translation aid is better integrated in professional environments which offer more direct training and support, sometimes related to higher education and academic institutions active in the field. Thus, our hypothesis is that in-house translators will adopt corpus use more readily than freelance translators. For this reason, these two respondent profiles are also specifically selected in our discussion of the reviewed surveys. For the discussion of the survey we have conducted for this paper (section 3) we also focus on the in-house versus freelance respondent profiles based on this hypothesis. For a resource such as that of corpora, which initially served academic research purposes, to be integrated in professional practice, it is advisable to integrate it in a CAT tool, which has become the basis of most professional translators' toolkits now.

First, we review in section 2.1 three surveys conducted within Europe (cf. Table 1), starting with the oldest one (MeLLANGE, 2006: UK and other respondents), then comparing the data to more recent European surveys from Spain (Gallego-Hernández, 2015: 526 respondents) and Switzerland (Picton, Fontanet, Maradan, & Pulitano, 2015: 202 respondents). We continue with data from the global surveys by Blancafort & Gornostay (2010) and by Zaretskaya *et al.* (2015, 2018) (cf. Table 2).

#### **2.1.2. Survey review**

A short survey in the framework of the MeLLANGE project (2006) researched, among other things, the use of conventional corpora in Europe. The largest group of respondents (567 or

56%) originated from the United Kingdom. In total, 1015 respondents completed the survey. Note that this survey was conducted among both student and professional translators, but the major part of the respondents (74%) were professionals (MeLLANGE, 2006, p. 4).

The survey established that overall 41.8% of all respondents used corpora when translating (MeLLANGE, 2006, p. 6). The professional translators' use of corpora even amounted to 44.2% in this survey (MeLLANGE, 2006, p. 10). It must be noted that English corpora in particular are widely available (Frankenberg-Garcia, 2015, p. 356). As the largest group of respondents from this survey is UK-based, this may influence the corpus user rate. However, this cannot be firmly established, as the respondents' answers were not subdivided by country. On the basis of this information, however, we will attempt to formulate an answer to the question of corpus use by professional translators and language pairs, by considering the results of our small-scale survey per language pair (section 3.4). On the basis of Frankenberg-Garcia's statement we hypothesize that corpus use will be more common among translators working with English in their language pair.

The types of corpora which were overall most frequently used were monolingual corpora in the target language (25.9%) and in the source language (22.8%) (MeLLANGE, 2006, p. 6). The main reason why respondents did not use corpora was because they were fully unacquainted with the concept of corpora (41%). But a great majority (84%) was interested in being provided with tailor-made corpora and tools for extracting terms from domain-specific corpora (83.4%). Furthermore, 85.9% of the respondents was also very keen on learning more about the potential of corpora (MeLLANGE, 2006, pp. 7-8), which showed that there was definitely a future interest for corpora among translators at that time already.

Other, more recent surveys conducted within Europe, for instance in Spain (Gallego-Hernández, 2015: 526 Spanish respondents) and Switzerland (Picton *et al.*, 2015: 202 respondents), confirmed these findings, with respectively nearly 50% and even 70% of respondents who use corpora sometimes, often or very often. In Picton *et al.*'s survey, the high percentage of in-house translators might also be a factor which impacts the corpus user rate positively, as companies and institutions may have more means and organizational structures which allow them to acquire new insights and stay up to date with new tools and possible support systems than freelance translators. In the Spanish survey multilingual corpora were the most frequently used corpus type (Gallego-Hernández, 2015, p. 381). The frequent use of multilingual corpora may be due to the relative unavailability of monolingual corpora for Spanish translators. The survey shows that 71% of the corpus users uses free online ready-made corpora. In addition, 66% compiles their own corpora (Gallego-Hernández, 2015, p. 381). For this compilation task Spanish translators may more often search for parallel texts (source texts and their translations), thus explaining the frequent use of multilingual corpora instead of monolingual corpora.

	location respondents	occupation respondents	overall corpus use
MeLLANGE (2006)	Europe, mostly UK (56%)	students (26%), professional translators and interpreters (74%)	41.8%
Gallégo-Hernández (2015)	Spain	in-house translators (14%), freelance translators (55%), project managers (3%), respondents with other sources of income (28%)	50%
Picton <i>et al.</i> (2015)	Switzerland	in-house translators (80%), freelance translators (18%), 'other' (1%)	70%

**Table 1.** Corpus use in Europe by respondent occupation (2005-2015)

The inverse magnitudes in Table 1 of in-house translators and freelance translators in Gallégo-Hernández' survey and in Picton *et al.*'s survey published in the same year (2015) combined with the frequencies for overall corpus use appear to confirm the above suggestion that in-house translators may have better access to corpora as a translation aid: Gallégo-Hernández' survey shows a 50% overall corpus use and includes only 14% in-house translators and 55% freelance translators, whereas Picton *et al.*'s survey shows a considerably higher corpus use, based on a clear majority of 80% in-house translators and only 18% freelance translators. Recent figures from a global survey by Zaretskaya *et al.* (2015, 2018) based on a clear majority of freelance translator respondents (91%) and only 6% in-house translator respondents also point to a very low adoption of corpora as a translation aid by freelance translators, thus confirming the above suggestion once more (cf. Table 2). Zaretskaya *et al.*'s study in 88 different countries worldwide queried the familiarity with translation technologies other than translation memories. It appeared that a very low rate of respondents used corpora (15%) and corpus tools (17%). Very few (freelance) translators use corpora and there are even fewer translators who compile their own corpora, because this task is simply too time-consuming (Zaretskaya *et al.*, 2015, p. 253). This was also already established by Bowker (2004) in her research on professional translators in Canada, when she stated that they often lack the time to set up corpora before they start their translation task (p. 240).

	occupation respondents	overall corpus use
Blancafort & Gornostay (2010)	in-house translators, freelance translators (30%), terminologists, language teachers and translator trainers, localization experts	50%
Zaretskaya <i>et al.</i> (2015, 2018)	students (2%), in-house translators (6%), freelance translators (91 %)	15%

**Table 2.** Corpus use worldwide by respondent occupation

In another fairly recent study with respondents worldwide, which surveyed respondents with more varied occupation profiles, including not only translators, but also terminologists, lan-

guage teachers, translator trainers and localization experts (cf. Table 2) 50% indicated that they collect corpora, particularly parallel and comparable corpora (Blancafort & Gornostay, 2010, p. 21), assumedly with the aim of using them as a translation resource.

Although the studies which have been reviewed in this section do not allow fully watertight comparison of respondents' profiles and corpus use, as they were not designed as replication studies with regard to respondent occupation, we see clear indications which point to higher adoption of corpora (not TM-based) by in-house translators than by freelance translators. As mentioned, in-house translators may benefit from the state-of-the-art and didactic environment offered by their companies. The same function is, of course, at the heart of higher education translation training environments, where corpora have proven to be useful resources (cf. Bowker, 1998, 1999; Kübler, 2003; Kübler, Mestivier & Pecman, 2018). This could also explain the 50% corpus collection rate in Blancafort and Gornostay's (2010) survey, as it includes language teachers and translation trainers among the respondents. But it must be added that few details are provided about the percentages of the respondents' occupations in their survey.

## 2.2. Surveying the use of CAT tools

### 2.2.1. Introduction and rationale

Following our overview of the use of corpora in translation practice in the recent past as a starting point for the integration of corpora in translators' daily professional practice, let us now have a look at the adoption and evolution of the tool which would allow this integration most naturally, viz., the use of CAT tools among different translator profiles.

In our survey the notion 'CAT tool(s)' includes translation memories (TMs) as a widely used component (cf. also Vandeghinste *et al.*, 2019, p. 3, on the core function of the TM for CAT tools). The following definition was provided for the respondents: "A CAT (computer-assisted translation) tool supports the translation process through software. Usually a CAT tool consists of several components, such as a translation memory (TM) and a terminology management system. A translation memory is a database which stores source segments and their translated target segments. Translation memories can be created from scratch in the CAT tool itself or translation memories can be imported in your CAT tool to retrieve segments immediately. Using translation memories makes the translation process faster and more efficient."<sup>4</sup>

The current section will review the following surveys on the aspect of CAT tool use: the survey by Picton *et al.* (2015) for Switzerland and the global surveys by Blancafort & Gornostay (2010) and Zaretskaya *et al.* (2015, 2018) which were also discussed with a focus on the use of corpora in section 2.1. In addition, the UK related surveys by Fulford & Granell-Zafra (2005) and Dillon & Fraser (2006) will be reviewed, as well as the global surveys by Lagoudaki (2006), Gough

<sup>4</sup> Cf. also the following description, which summarizes the main functions of CAT tools well: "A CAT tool is a computer program that helps to translate text documents more efficiently through four main functions: [i] A CAT tool segments the text to be translated in segments (sentences) and presents the segments in a convenient way, to make translating easier and faster. [...]", [ii] The translation of each segment is saved together with the source text. Source text and translation will always be treated and presented as a translation unit (TU) [...], [iii] The main function of a CAT tool is to save the translation units in a database, called translation memory (TM), so that they can be re-used for any other text, or even in the same text. Through special 'fuzzy search' features the search functions of CAT tools even find segments which do not match 100 %. This saves a lot of time and effort and helps to make a coherent and consistent translation, [iv] The fourth basic function of a CAT tool is the automatic look-up in terminology databases, and the automatic display and insertion of the search results. (formally adapted from <http://www.metatexis.com/cat.htm>). Cf. also Vandeghinste *et al.* (2019, p. 3) on improved fuzzy matching).

(2011), Allard (2012) and Van den Bergh *et al.* (2015). Apart from the focus on prevalence of general CAT tool use, attention will also again be devoted to respondent profiles (in-house versus freelance especially), and recent evolution of CAT tool adoption.

### 2.2.2. Survey review

In a survey by Fulford and Granell-Zafra (2005) reporting on the use of translation technologies among freelance translators (89% of respondents) and in-house translators (11% of respondents) in the United Kingdom only 28% indicated that they used CAT tools, such as *Trados*, *SDLX*, *Transit* and *Déjà Vu* (p. 10). But another survey around the same time targeting the same freelance and in-house translation profession profiles in the United Kingdom (Dillon & Fraser, 2006) reported a total of 52% of TM users (p. 72). The respondent profiles for Dillon & Fraser's survey (2006) included similar numbers of in-house translators and freelance translators as Fulford and Granell-Zafra's (2005) survey, with respectively (11%) and (85%); some respondents belonged to both the in-house and freelance category (4%) (p. 72) (cf. Table 3). We should note, however, that Fulford and Granell-Zafra's lower CAT tool adoption results related to the freelance translators only (p. 8). Picton *et al.*'s more recent survey (2015), whose respondents are mainly in-house translators, shows a higher percentage of CAT tool users (82%).

	location respondents	occupation respondents	overall CAT tool use
Fulford & Granell-Zafra (2005)	UK	in-house translators (11%)*, freelance translators (89%)  *Survey results relate to freelance translators only	28%
Dillon & Fraser (2006)	UK	in-house translators (11%), freelance translators (85%), both freelance and in-house translators (4%)	52%
Picton <i>et al.</i> (2015)	Switzerland	in-house translators (80%), freelance translators (18%), 'other' (1%)	82%

**Table 3.** CAT tool use in Europe by respondent occupation (2005-2015)

Apart from an evolution over time, the survey results on the adoption of CAT tools discussed above may reflect freelance translators' hesitation to implement new technologies, such as CAT tools in the recent past, compared to in-house translators, whose adoption of new translation technologies might be connected to their agencies' direct professional support. This reflects the suggested conclusion in the previous section (2.1) relating to the adoption of corpora as a translation aid among freelance translators versus in-house translators with relevant figures from especially more recent findings (Gallégo-Hernández, 2015; Picton *et al.*, 2015; Zaretskaya *et al.*, 2015, 2018). The adoption of CAT tools by professional translators precedes a potential adoption of corpora as a possible translation aid; and a pattern based on translators' profiles may shine through here.

Apart from an evolution towards increased CAT tool use (incl. TM) over the years other factors also play a role. Let us have a look at the global surveys on the use of CAT tools. Around the same time as Fulford & Granell-Zafra's and Dillon and Fraser's UK surveys Lagoudaki (2006)

established in her worldwide survey that 82.5% of the respondents used a TM system. The high rate of TM system users in this case is due to the fact that most respondents specialized in technical texts, which in most cases contain “a great amount of terminology, standard expressions [...] and a high degree of internal repetition; therefore, the possibilities for content re-use are many” (Lagoudaki, 2006, p. 12). Blancafort and Gornostay’s (2010) study similarly shows high CAT tool (incl. TM) adoption (74%) for their mostly technical translation profiled respondents (p.6). The studies by Gough for global in-house and freelance translators (2011, p. 203) and Allard (2012, p. 399) showed increasing frequencies for CAT tool use, respectively over 80% and 90.8%. The latter high rate may be explained by the fact that CAT tool use was one of the conditions for participation in the survey, which focused on terminology management, however.

In the following years, global CAT tool utilization rates consistently reached at least 75% (Zaretskaya *et al.*, 2015, 2018; Van den Bergh *et al.*, 2015). The respondents’ occupations in the different surveys do not greatly influence the CAT tool use rate at this later stage in the evolution of CAT tool use by professional translators, although we do see the highest adoption rate (90.8%) in the survey which lists the highest proportion of in-house translators in Table 4.

	occupation respondents	overall CAT tool use
Lagoudaki (2006)	freelance translators (73%); company owners (8%); company employees (19%)	82.5%
Blancafort & Gornostay (2010)	in-house translators, freelance translators (30%), terminologists, language teachers and translator trainers, localization experts	74%
Gough (2011)	mostly freelance translators	> 80%
Allard (2012)	in-house translators (44%), freelance translators (56%)	90.8%
Zaretskaya <i>et al.</i> (2015, 2018)	students (2%), in-house translators (6%), freelance translators (91%)	76%
Van den Bergh <i>et al.</i> (2015)	in-house translators (24%), freelance translators (72%), terminologists (11%), interpreters (10%), project managers (7%), post-editors (7%)	> 75%

**Table 4.** CAT tool use worldwide by respondent occupation (2006-2015)

The most common, albeit vague, reason why respondents indicated they do not use a TM system in Lagoudaki (2006) was unsuitability of TMs for their work. Other reasons were the high purchase cost, dissatisfactory TM trial experience and the lack of benefit. Surprisingly, also 16% of the respondents at the time had not had the opportunity yet to learn how to use the TM tool which they possessed. However, on a promising note, 71% of the non-users intended to try out or buy a TM system (Lagoudaki, 2006, p. 14). The motivations to use a TM were firstly economizing time, and secondly terminological consistency and quality increase. According to the respondents of this study from 2006 (p. 18), TM systems also save costs and it is the best approach for exchanging glossaries, TM databases, etc. These tasks have not changed

much over the years, as confirmed by the survey by Zaretskaya *et al.* (2015), which also lists the following tasks executed with TMs in addition to translation: “consistency check, import of TM, terminology management, completeness check, analysis for invoicing, terminology quality assessment, terminology extraction, among others” (p. 251).

On the question how respondents had acquired CAT tool skills Van den Bergh *et al.* (2015) established that most global users were self-taught (p. 115). Only 29.45% of the respondents in Zaretskaya *et al.*'s (2015) survey had followed specialized courses on CAT tools (p. 249). But in the earlier survey by Lagoudaki (2006) almost half of the respondents stated that they had received training on TM technologies, in various forms, i.e., short courses or seminars, company trainings, colleagues, etc. (p. 16). However, these differences in CAT tool training might be attributed once again to the difference in the different surveys' respondent profiles: whereas Zaretskaya *et al.*'s (2015, 2018) survey includes only students, in-house and freelance translators, Lagoudaki (2006) surveyed translators, project managers, reviewers, subtitlers, terminologists, interpreters, other translation professionals (DTP specialists, graphic designers, web authors, ...).

In conclusion – and excluding confounding factors (technical translation profiles in Lagoudaki, 2006 and CAT tool use as a condition for survey participation in Allard, 2012) – there are clear indications on the basis of the data in Tables 3 and 4 discussed above that freelance translators have caught up on CAT tool adoption over the years, starting around 2005 to the more recent surveys in 2015.

## 2.3. Surveying the use of terminology management systems

### 2.3.1. Introduction and rationale

Before turning our attention to some additional aspects of the use of CAT tools, terminology management systems (TMS) and corpora on the basis of our survey in comparison with another Belgian survey in section 3 we finalize our review of recent surveys with that part of translators' work which was systematized earliest, albeit first using personally devised systems, viz., that of terminology management.

In our survey *terminology management* is defined for the respondents as “the collection, storage and retrieval of terms. Terminology management can be performed using standard tools such as Microsoft Word or Excel. Specific *terminology management systems* are also available, which can be used stand-alone or can be integrated in a CAT tool. An example of a terminology management system is SDL MultiTerm, which can be integrated in SDL Trados.” A termbase is defined in the survey as “A database containing terminology”.<sup>5</sup>

In addition to the surveys reviewed above from Blancafort & Gornostay (2010), Allard (2012) and Zaretskaya *et al.* (2015, 2018), we also review earlier data from Zielinski & Ramirez (2005) and more recent data from Steurs & van der Lek-Ciudin (2016) in the current section.

### 2.3.2. Survey review

Over the years survey research has shown that terminology management has always been part of the activities of language professionals worldwide (Allard, 2012; Blancafort & Gornostay, 2010; Steurs & van der Lek-Ciudin, 2016; Zaretskaya *et al.*, 2015, 2018; Zielinski &

<sup>5</sup> Cf. also Bowker (2015, p. 310-311) on the evolution from personal terminology collections in word processors, spread sheets, etc. in the past to specially designed terminology management systems with term bases which allow integration with other tools such as translation memories (TMs) and term extractors. Cf. also Popiolek's (2015, p. 347) description of the key components of terminology management infrastructure as a database which allows storage and editing, searching and retrieval of terminology, ideally in an automated manner, among other aspects.

Ramirez, 2005), although some respondents in Allard's study indicated "that other resources such as the World Wide Web, existing glossaries and dictionaries, or online corpora meet their terminological needs" (2012, p. 125).<sup>6</sup> Some respondents lacked knowledge, thought it was not their responsibility, found it was too time-consuming or did not see the added value of terminology management (Steurs & van der Lek-Ciudin, 2016, p. 13). Respondents involved in terminology management have mostly used TMSs which were integrated into their main CAT tool in the recent past (Allard, 2012, p. 125; Steurs & van der Lek-Ciudin, 2016, p. 18). As with the market leading position of *Trados Studio* for translation memory systems (TMs) (cf. global surveys by Lagoudaki, 2006; Blancafort & Gornostay, 2010; Allard, 2012; Van den Bergh *et al.*, 2015) *Trados MultiTerm* – which can function as a stand-alone tool or can be integrated in *SDL Trados* – also ranks first as the most popular brand for terminology management systems. (Blancafort & Gornostay, 2010, p. 11; Steurs, De Wachter, & De Malsche, 2015, p. 228; Zielinski & Ramirez, 2005, p. 3).

But terminology is also still recorded in spreadsheets (e.g., *Excel*) and word processors (e.g., *Word*) (Allard, 2012, p. 126; Steurs & van der Lek-Ciudin, 2016, p. 16). Not all respondents recorded terminology in their integrated terminology management system, as some found it too complex, had received too little training or the system was not suited to their needs (Allard, 2012, p. 125-126). The sources used by respondents to collect terminology are glossaries, dictionaries and databases, as well as client resources according to Blancafort & Gornostay's study (2010, p. 13).

	occupation respondents	overall TMS use
Zielinski & Ramirez (2005)	freelance translators (65%), in-house translators (28%), both freelance and in-house translators (5%)	91%
Blancafort & Gornostay (2010)	in-house translators, freelance translators, terminologists, language teachers and translator trainers, localization experts	56%
Allard (2012)	in-house translators (44%), freelance translators (56%)	86.4%
Zaretskaya <i>et al.</i> (2015, 2018)	students (2%), in-house translators (6%), freelance translators (91 %)	58%
Steurs & van der Lek-Ciudin (2016)	in-house translators (25%), freelance translators (73%), revisors/editors (24%), terminologists (11%), interpreters (10%), post-editors (6%)	74%

**Table 5.** Terminology management systems worldwide by respondent occupation (2005-2016)

When we consider TMS use among freelance translators compared to in-house translators in the surveys referenced in Table 5, which provide frequencies for these user profiles, the earlier study by Zielinski & Ramirez (2005) shows a high adoption of TMS among a respondent population with a majority of freelance translators (65%). One explanation which might be suggested is that some freelance translators may have been more dependent on TMS which were not integrated in a CAT tool at a time when the use of a CAT tool with a translation memory was still not pervasive among most translators prior to 2005. The survey Allard (2012) conducted

<sup>6</sup> The specific terminological needs are not directly specified in Allard (2012).

somewhat later included a more balanced division of freelance translators (56%) versus in-house translators (44%) and shows a similar high adoption of TMS. We should note, however, that Allard's (2012) survey was aimed particularly at CAT tool users. As CAT tools frequently offer built-in TMS systems, this might perhaps explain the high TMS user rate (86.4%) among the survey respondents in this case. Conversely, relying on translation memory (TM) segments while translating may perhaps appear to reduce the need to consult terminological databases in a translator who is working against the clock; this might reduce the *perceived* need to consult a TMS as frequently as a translator might have done who did not rely on the TM segments offered by a CAT tool. This might perhaps explain the fairly low rate (58%) of TMS use in Zaretskaya *et al.*'s more recent study (2015, 2018), when CAT tools had already become far more widespread than was the case around the turn of the century. Steurs & van der Lek-Ciudin's (2016) recent survey, which includes a fair majority of freelance translators, but also other diverse profiles, however, shows a higher adoption of TMS (74%), but still a much lower rate than the 91% TMS adoption in Zielinski & Ramirez' (2005) earlier study.<sup>7</sup> A conclusion which may be suggested here is that somewhat more contextualized input for translators (i.e., in the form of TM segments) than that of more isolated terms may reduce the perceived need for explicit TMS consultation among professional translators. In our survey conducted in 2017-2018 we found that translation is the most frequently executed task with CAT tools (79% of respondents), followed by the import of translation memories (53%) and terminology management (52%) (cf. section 3).

### 3. The current situation among Belgian and Dutch translation professionals

#### 3.1. Introduction to our survey, general findings and the CBTI-BKVT survey

By their very nature TM segments offer fairly limited context. Therefore, we also wanted to focus on specific aspects of the use of corpora as a translation resource (cf. section 3.4). In the current section we report on the findings of a small-scaled survey which we conducted on CAT tool use, terminology management and especially corpus use among translation professionals in Belgium and the Netherlands. Apart from general comparison in sections 3.1 and 3.2 with the previous surveys reviewed above we also added two additional specific points of interest in sections 3.3 and 3.4 to the findings provided by the earlier, reviewed surveys, viz., the use of CAT tools, terminology management systems and also corpora by job role, language combination and text domain.

Our interest in the job role data was inspired by the approximative conclusions which were drawn on the basis of the reported proportions of freelance versus in-house translators in the reviewed surveys. On this point we hypothesized a higher adoption of these three types of translation aids among in-house translators (section 3.3). With respect to corpora we focussed on language combinations: as there is a longer tradition of corpus use for English and more English corpora have been available for English since the development of corpora and corpus linguistics our hypothesis was a higher adoption of the use of corpora by translators working with English in their language combination (section 3.4). Finally, we also considered the use of corpora in function of the translators' main text domain in section 3.4, as we expected a possibly higher uptake of the use of corpora for more specialized text domains.

Between March 2017 and February 2018 we gathered 116 surveys in total.<sup>8</sup> We received 101 surveys for which all questions were answered. Our results are based on those 101 fully com-

<sup>7</sup> Blancafort & Gornostay's (2010) study was not added to the discussion here, as (i) their survey does not provide percentages for the different respondent profiles and (ii) it also included occupational profiles which assumedly use TMSs more frequently, such as terminologists, translator trainers and localization experts.

<sup>8</sup> The initial goal was to collect between 100 and 200 fully completed surveys.

pleted surveys. We first collected general profiling information, such as age group, main qualifications, job role (freelance versus in-house translators), years of experience, language combination and specialization (cf. Appendix 1 for the complete survey, containing 25 questions). Respondents were sought by contacting the *Belgian Chamber of Translators and Interpreters* (CBTI-BKVT) in 2017. Survey invitations were also e-mailed to translation agencies and freelance translators were located through the Yellow Pages (Belgium) and the website of the *Vereniging van Vertaalbureaus in Nederland (VViN)* (*Netherlands Association of Translation Companies*) and <https://www.vertaalbureau-info.nl/azindex.php> for the Netherlands. Calls for respondents were also launched on Facebook.

In January 2018 the *Belgian Chamber of Translators and Interpreters* (CBTI-BKVT) opened a survey to its own members for a period of forty days throughout all regions of Belgium. The CBTI-BKVT survey mainly presents information on current translation rates and related aspects such as income and revenue. In addition, it also queried aspects such as workload satisfaction, etc. (CBTI-BKVT, 2018). This market-oriented survey collected 439 surveys, but not all respondents filled out all questions, as there was no obligation to do so. Thus, the sample size varies for the different questions of the CBTI-BKVT survey. Apart from questions related to financial aspects some other questions overlap with our survey. Our survey focusses on the use of CAT tools and terminology management systems and adds the category of corpora as a translation aid. In addition, we collected separate data for the job roles of freelance translators and in-house translators, with the aim of comparing the potential impact of a translator's direct professional environment on the use of these translation aids.

Although it is difficult to compare the demographic set-up of the CBTI-BKVT survey and our survey in terms of respondents' ages because of different age classifications in the survey questions, the CBTI-BKVT survey appears to report on a somewhat overall younger population than our survey, which includes fairly similar figures for all younger age groups, but a slightly larger group of respondents between 51 and 60 years of age (27%), who have over 20 years of experience. As in the CBTI-BKVT survey the largest group of our respondents have more than 20 years of translator experience. The other respondents have similar experience profiles in both surveys, apart from a somewhat larger group of the least experienced in the CBTI-BKVT survey.

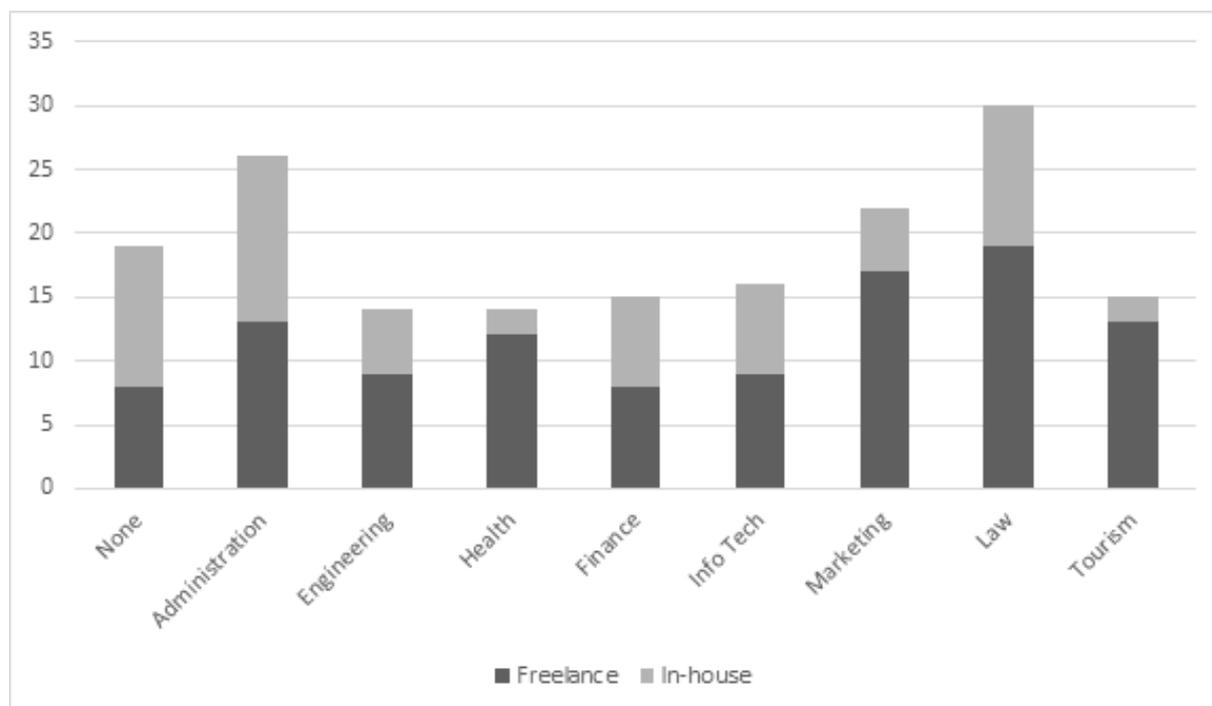
The largest group of respondents in our survey has a master degree in translation (41%); other qualification profiles in our survey are a master degree in languages (19%), but also a bachelor degrees in translation (17%), whereas the CBTI-respondents virtually all have master degree, mostly in translation (70% of respondents). Like the CBTI-BKVT respondents most of our respondents work as freelance translators, with 77% and 64% respectively. English/Dutch is the most frequent language combination among our respondents (72%), followed by French/Dutch (57%) and German/Dutch (35%). In this respect our survey, which included Dutch translation professionals as well, differs from the Belgian CBTI-BKVT survey, as our survey focussed on language combinations including Dutch, with French, English, German and Spanish in both translation directions.

For a consideration of the text domains and specializations we listed a number of domain options in our survey (cf. Table 6 below). The respondent group of freelance translators translate legal texts most often (29.2%). The smaller respondent group of in-house translators also indicated legal texts as one of their main text types for translation (30.6%), together with administrative texts (36.1%) (cf. Table 6). This is also one of the more frequent text types for the freelance translators (20%), together with texts related to health (18.4%), tourism (20%) and especially marketing (26.2%). The domains of health and tourism are not frequent among the in-house translators in our survey, however (both 5.6%).

	None	Administration	Engineering	Health	Finance	Information Technologies	Marketing	Law	Tourism
% of Freelance translators	12.3%	20.0%	13.8%	18.4%	12.3%	13.8%	26.2%	29.2%	20.0%
% of In-house translators	30.6%	36.1%	13.9%	5.6%	19.4%	19.4%	13.9%	30.6%	5.6%

**Table 6.** Text domains: relative frequencies of freelance translators and in-house translators in the current 2017-2018 survey

When we consider the absolute frequencies of the freelance respondents and the smaller group of in-house respondents of our survey (cf. Figure 1 below), law and administration again show up as overall dominant domains. The CBTI-BKVT survey lists domain options related to those in this paper’s survey, but based on a different a division and a different querying methodology concerning text domains, so that a full comparison is not possible. But for the question on the respondents’ *first* specialization, law is also the most frequent domain (respondents’ *first* specialization (21%) in the CBTI-BKVT survey (p.14). Marketing was not frequently indicated as the CBTI-BKVT respondents’ first specialization.<sup>9</sup>



**Figure 1.** Text domains: absolute numbers of freelance translators and in-house translators in the current 2017-2018 survey

In addition to the listed domain options respondents also frequently added other text types or general subjects, such as mobility and transport, gastronomy, technical, automotive, arts and also a small proportion for literary translation (5%).

<sup>9</sup> Considerations on translation of marketing related texts as ‘transcreation’, a job role which differs from the traditional translator’s job role, would lead us beyond the scope of the current article.

### 3.2. The use of CAT tools, terminology management systems and corpora in our survey (2017-2018): general findings and comparison with the CBTI-BKVT survey

We found that a vast majority of the respondents in our survey (80%) uses CAT tools. This is in line with previous surveys conducted in Europe (Picton *et al.*, 2015) and worldwide (Allard, 2012; Gough, 2011; Van den Bergh *et al.*, 2015; Zaretskaya *et al.*, 2015). This result is also in line with the 77% CAT tool use among 421 respondents in the Belgian CBTI-BKVT (2018) market survey report, which includes three basic questions concerning the use of CAT tools among all translator profiles, summarized in three graphs (p. 30-31), viz., (i) “Do you use CAT tools [...] or similar tools when translating?”, (ii) “Why don’t you use CAT tools” and (iii) Which CAT tools do you use?”<sup>10</sup> (cf. below for comparison with our findings on the last two questions).

In our survey 82 out of 101 respondents indicated that they are specialized translators. The great majority of these specialized translators (64) also uses CAT tools, which corresponds to the finding that a high TM usage rate relates to (technical) translation specialization (Blancafort & Gornostay, 2010, p. 5; Lagoudaki, 2006, p. 12).

Concerning the question which CAT tools professional translators use, *SDL Trados* has maintained a strong global position within the translation industry over the years. This is also shown by the fact that in our survey, in correspondence to the surveys previously discussed (Allard, 2012; Blancafort & Gornostay, 2010; Lagoudaki, 2006; Van den Bergh *et al.*, 2015), *SDL Trados* still ranks first (72%). *MemoQ* ranks second in our survey with 30%, reflecting the CBTI-BKVT (2018) findings (p. 31). Over half of the respondents of our survey received CAT tool training (56), many in the form of external training (34 respondents specifically referred to workshops, seminars, webinars, etc.; cf. Lagoudaki, 2006). Not surprisingly, translating is the most frequent activity with CAT tools (79% of respondents), followed by the import of translation memories (53%) and terminology management (52%). Of the 20 respondents who do not use CAT tools, 9 state they do not need them (45%), possibly because of the domains they specialise in, e.g., literature, audiovisual translation, history and art. These are domains which are generally not characterized by repetition and are therefore less suitable for CAT tool use. However, 6 of the non-CAT tool users (30%) plan to use CAT tools in the future. 5 of those are less experienced translators (1-5 years of experience). These findings reflect the CBTI-BKVT findings, where 60% of those who do not use CAT tools indicated that they do not need them, but – more relevantly perhaps – as much as 26% also stated that they do not know how to use CAT tools. (p. 30). We can also assume that over the years a proportion of those translators who reported an *intention* to start using a CAT tool has actually done so, as the indication of such an intention (compared to actual use) was higher in past surveys (Lagoudaki, 2006) than in ours.

Concerning terminology management, 65% of the respondents make use of integrated or stand-alone terminology management systems mostly. As expected, based on the CAT tool results where *SDL Trados* appears to be most popular, *SDL MultiTerm* also ranks first for terminology management (70%), in line with previous surveys (Allard, 2012; Steurs & van der Lek-Ciudin, 2016). When asked how respondents engaging in terminology management obtain terminology, they mostly indicate that they compile their own termbases or terminology lists based on source texts and their translations for future use (56%).

With 48 versus 53 respondents respectively the difference between corpus versus non-corpus use is much smaller than the difference between CAT tool versus non-CAT tool use (81 versus 20). But, despite providing written definitions in the survey as well as in most of the e-mailed survey invitations to define what we considered to be a TM on the one hand and a corpus on

<sup>10</sup> The CBTI-BKVT (2018) survey next reports on the use of MTPE (machine translation post-editing (p. 31-32) among professional translators.

the other, this distinction did create confusion for some of the respondents, as they stated they used TMs to build corpora.<sup>11</sup> TMs provide matching content for segments from the source text which is to be translated. This content is harvested from segments in the TM; hence context beyond the segments is not taken into account. In other words, TMs can be described as repositories of which all the sentences it contains are out of context (Bowker, 2006, p. 179). A corpus is a running text; hence relevant textual context beyond the sentence (or segment) is available. It must be added here that often corpus query systems allow the user to retrieve limited sections (a number of sentences). This context can sometimes be expanded. Overall, corpus use provides more context than TM use does (cf. Tatu, 2011; Bowker & Barlow, 2008). Due to the confusion stated above, the real number of corpus users may be lower than shown in our survey.

When we compare how the findings from our survey for CAT tool use and corpus use differ from previous surveys, we see that Blancafort & Gornostay (2010) (who surveyed varied respondent profiles but did not provide clear indications for the proportions of the respondent profiles in their survey) report findings very similar to ours, with 74% CAT tool use and 50% corpus use (cp. also 50% corpus use in Gallégo-Hernández [mainly multilingual], 2015 and 41.8% MeLLange, 2006). Interestingly, however, the survey which has a high proportion of in-house translators (Picton et al., 2015) reports a very similar frequency of CAT tool usage (82%) compared to our findings, but a considerably higher frequency of corpus use (70%) than our finding for a respondent group with only 36% in-house translators. And Zaretskaya *et al.* (2015, 2018), whose main respondent profile is that of freelance translators (91%), report very low corpus use (15%) (in addition to a 76% CAT tool use, which is similar to the other findings reported here for CAT tool use). From this we might tentatively conclude that corpora, as the lesser known potential translation aid, finds easier adoption in the recent surveys among in-house translators than freelance translators. This apparent conclusion was taken as the basis for a hypothesis along these lines tested for the data from our survey by dividing survey responses for both translator profiles, as will be discussed in section 3.3.

Parallel corpora (source texts and their corresponding translations) are clearly the most popular corpus type (81%) among the respondents in our survey. 38% of the respondents also build their own corpora. However, nearly half of these do-it-yourself corpus compilers mention translation memories for corpus building. This indicates once more the blurred line between a translation memory as a feature of a CAT tool and a corpus. Other, genuine corpus building methods mentioned by our respondents are *MemoQ* and *MultiTrans*, which are CAT tools incorporating corpus building features (cf. *infra*). 53% of the non-corpus compilers indicate not being interested in building their own corpora. Finally, respondents who do not use corpora mainly do not do so because they are not familiar with them (28 out of 53). This confirms the findings of previous surveys on corpus use (MeLLANGE, 2006; Zaretskaya *et al.*, 2015). 34 out

---

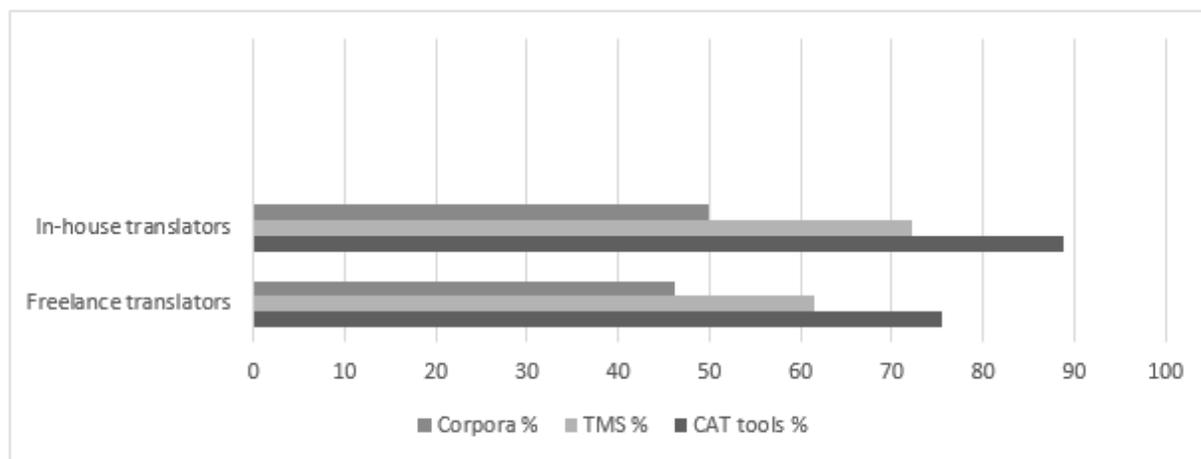
<sup>11</sup> The following succinct definitions of a corpus and different types of corpora were provided to the respondents in the survey text (cf. Appendix 1): “A corpus in this survey refers to a collection of written texts which is used for searching equivalent terms while translating. A corpus which contains original texts in one particular language is called a monolingual corpus. A corpus including collections of texts with similar content in different languages or language varieties (e.g., British English, American English, Australian English) is called a comparable corpus. A comparable corpus does **not** consist of original texts along with their corresponding translations (cp. parallel corpus). Therefore, the (linguistic) content is not influenced by translation processes. A parallel corpus consists of original texts in one particular language, along with their translations in a different language. Examples of parallel corpora on the web are Linguee and Reverso Context, which are ready-made corpora. However, translators can also compile their own corpora.”

of the 53 non-users in our survey are also not planning on using corpora in the future. This shows once more that it is necessary to raise corpus awareness among translators (cf. also Frankenberg-Garcia, 2015).

### 3.3. CAT tools, terminology management systems and corpora: A comparison of freelance and in-house translators' work practice

In addition to the more general questions discussed in section 3.2, we also studied the use of CAT tools, terminology management systems and corpora in relation to job role on the basis of our survey data.

When we compare the use of the three types of translation aids in Figure 2 for Belgium and the Netherlands by freelance versus in-house translators in our survey, we see a considerably larger adoption of both CAT tool support and terminology management systems, with respectively 13.5% and 10.7% higher uptake by in-house translators in our survey. However, contrary to the hypothesis formulated in section 2.2 that corpora find easier adoption as a translation aid among in-house translators than their freelance colleagues, based mainly on findings from Picton *et al.* (2015) and Zaretskaya *et al.* (2015, 2018) and their respective respondent profiles, the use of corpora is not only less popular than the use of CAT tools among our respondents, but also has very similar adoption in *both* job roles in our small scale survey: half of the in-house translators and 46.2 % of the freelance translators stated that they use corpora to support their work. Assumedly, the fact that corpora are not integrated in the translation workflow systems, unlike CAT tools and terminology management systems, affects both categories of translators in a similar way.

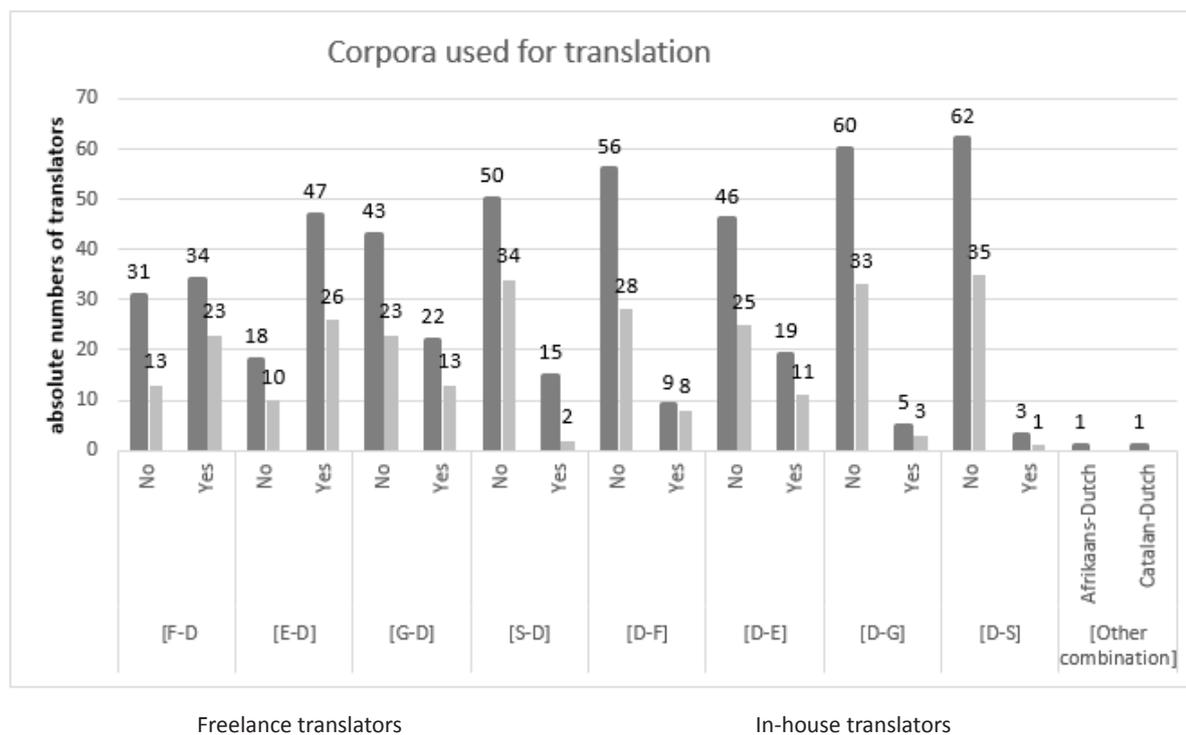


**Figure 2.** Use of CAT tools, terminology management systems and corpora by Belgian and Dutch freelance and in-house translators in the current survey (2017-2018)

### 3.4. Use of corpora by language combination and types of corpora

Apart from the general findings we especially wanted to consider the use of corpora as a translation aid for different language combinations, as we expected a higher uptake of the use of corpora by translators working with English in their language combination, in view of the longer and more extensive availability of English corpora compared to other languages (cf. section 2.1.2 with reference to MeLLANGE, 2006 and Frankenberg-Garcia, 2015). This hypothesis is confirmed by the findings of our small-scale survey. We see a considerably higher uptake of corpora among both in-house and freelance translators for the language combination English/Dutch compared to all other language combinations in our survey (French/Dutch,

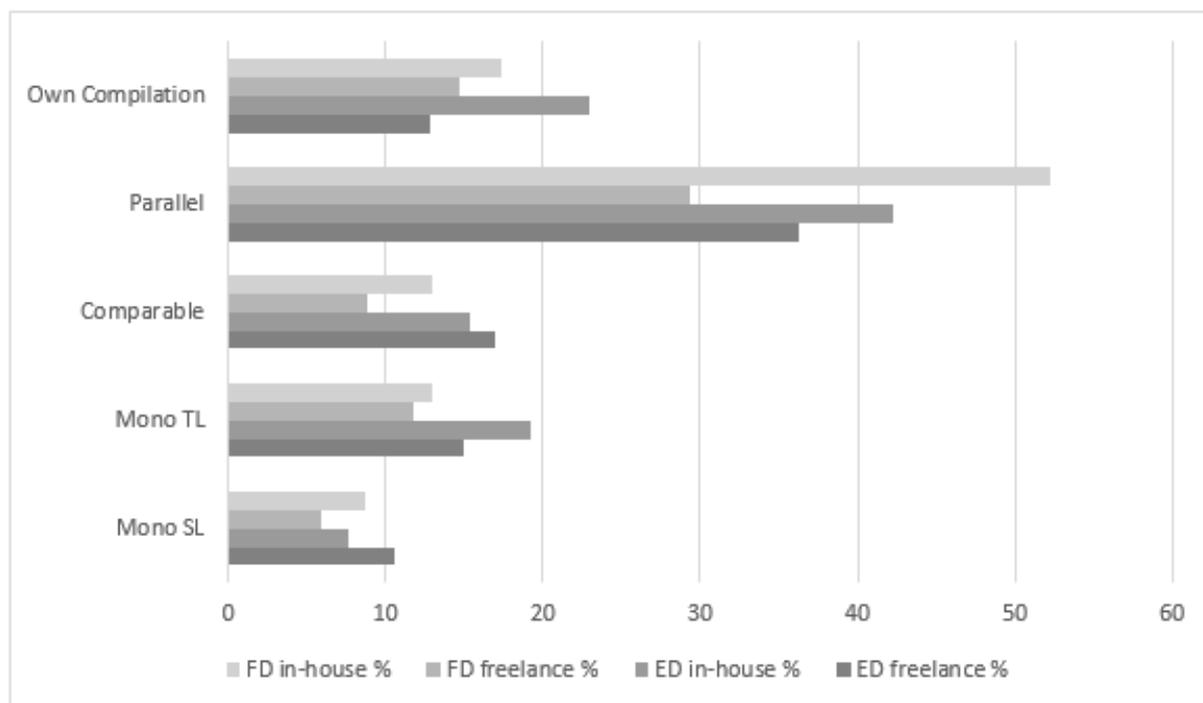
English/Dutch, German/Dutch, Spanish/Dutch and all the reverse combinations). Interestingly, however, this is not the case for Dutch to English translators in our survey, where many more respondents indicated that they do not use corpora (cf. Figure 3).



**Figure 3.** Corpus use per language combination in the current survey (2017-2018)

We also found a somewhat higher uptake of the use of corpora for translations from French into Dutch by both freelance and in-house translators than the other language pairs. But, as for English and Dutch, the opposite is clearly true for the reverse translation direction (cf. Figure 3). Whereas our respondents indicated that they used corpora clearly more often for English to Dutch translations, and somewhat more often for French to Dutch translations than for the other language pairs, corpora are not often used for translations from Dutch (probably most of our respondents' mother tongue) into English or French in our survey.

Let us now consider which types of corpora are used and compare with the MeLLANGE survey (2006) (cf. section 2.1.2), where most frequently monolingual corpora in the target language (25.9%) and in the source language (22.8%) were used. In terms of relative frequencies for the translation pairs English/Dutch and French/Dutch our respondents mostly indicated parallel corpora. This points to the traditional use of TMs, and as mentioned before, might perhaps overlap with it in respondents' interpretation, despite the definitions of corpora provided in the survey. Parallel corpora appear to be used especially by the French/Dutch in-house translators (cf. Figure 4). This may be related to the two main official languages of Belgium, which has resulted in the availability of a wealth of (official) bilingual text materials. But we also see a higher percentage of the use of parallel corpora among the English/Dutch in-house translators compared to the English/Dutch freelance translators in Figure 4. Overall, the group of French/Dutch in-house translators makes more frequent use of all types of corpora distinguished in our survey than their French/Dutch freelance colleagues (cf. Figure 4). The English/Dutch in-house translators make more use of self-compiled, parallel and monolingual corpora in the target language, but the English/Dutch freelance translators make slightly more use of monolingual corpora in the source language and comparable corpora.



**Figure 4.** Type of corpus for freelance and in-house translators with language pairs English/Dutch and French/Dutch<sup>12</sup>

In order to assess the use of corpora as a translation resource in its own right, unrelated to the more traditional recourse to parallel corpora (incl. possibly ready-made corpora readily available on the internet such as Linguee and Reverso Context) and TMs, we also considered total frequencies *excluding* parallel and self-compiled corpora. The total relative frequencies for monolingual corpora in the source text and target text languages and comparable corpora amounts to approximately 42 % for both freelance and in-house English/Dutch translators, but drops to 26.5% and 34.7% for their French/Dutch freelance and in-house colleagues respectively. When we consider recourse to self-compiled corpora the English/Dutch in-house translators clearly stand out (cf. Figure 4).

A comparison of the degree of corpus use by translated text domain in our survey did not yield any considerable differences, apart from a slightly higher figure of corpus use for legal texts by both freelance and in-house translators, and very similar proportions of use and non-use for administrative texts by in-house translators only.

Obviously, the small-scale survey can only point to possible tendencies in terms of the use of (i) corpora for different language pairs in a given language context (e.g., Belgium) and (ii) different types of corpora as translation resources. But it may point to some categories and distinctions worth investigating further, based on clear definitions of corpora and corpus types on the basis of larger data sets.

<sup>12</sup> The frequencies are relative to the following totals: 47 English/Dutch freelance, 26 English/Dutch in-house, 34 French/Dutch freelance and 23 French/Dutch in-house translators.

#### 4. The future: towards an integration of corpora in CAT tools?

The survey results discussed earlier show that CAT tools and their TM component are very frequently used in current professional translation practice. The use of TMs entails both advantages and disadvantages. Bowker (2005, p. 15) and Leblanc (2013, p. 6) state that they improve translation quality by increasing consistency. When several translators work on the same translation project, TMs can be shared, which has a positive influence on translation consistency for collaborative translation projects as well (Bowker, 2005, p. 15). However, the use of TMs does not always benefit translation quality. In some case TM output is based on different TMs provided by different translators; if the TM is not properly maintained this negatively impacts consistency. Moorkens' (2012) study about TM quality also points to faulty TM maintenance as one of the causes of inconsistency. This confirms that TM translations are often inconsistent (Moorkens, 2012, p. 210). As translation quality depends on the content of the TM, poor quality TMs will also negatively impact overall translation quality. Translators may lose time they would have saved using a high quality TM when they need to correct translations generated by poor quality TM content. The fact that TMs do not typically store full running texts, but rather isolated segments without any context, contrary to corpora, may also negatively impact the quality of translations (Bowker, 2005, p. 15).<sup>13</sup> In addition, quality issues may arise when the TM is no longer up-to-date due to changes in terminology, or a proposed translation might not fit the context (Bowker, 2005, p. 19). Another disadvantage is the time and effort needed to learn how to master a TM system (Fernández-Parra, 2010, p. 3; García, 2006, p. 98).

To address some of the drawbacks of TM translation, such as decreased consistency and de-contextualization, corpora could be incorporated in translation practice. However, in order for corpus use and corpus compilation/search tools to become popular among professional translators they probably ought to be integrated in CAT tools (incl. TMs) and translators' current workflow. Such an integration, which has already taken place with regard to terminology management systems in CAT tools, requires interaction with web search engines in order "to search, retrieve and morphologically annotate corpora based on user specifications" (Bernardini & Castagnoli, 2008, p. 52). Full and fuzzy matches would be searched automatically in gold standard TMs, while hypotheses could be tested and developed by "manual concordancing of comparable and parallel texts" (without any contribution from TMs) (Bernardini & Castagnoli, 2008, p. 52).<sup>14,15</sup> The integration of corpora and corpus features, such as concordancing (Bowker & Barlow, 2008, p. 52; Tatu, 2011, p. 173) (finding a search word or pattern in context), into TMs leads to so-called hybrid tools (Bowker, 2004, p. 223; Bowker & Barlow, 2008, p. 20). In Table 7 we compiled information from Tatu (2011) and Bowker & Barlow (2008) to compare TMs and corpora / corpus features (such as bilingual concordancing); Table 7 also shows in which ways the implementation of corpora / corpus features in TMs can be beneficial.

<sup>13</sup> SDL Trados provides a Context match function, however, where the previous/next segment is retained to flag matches as context match, so that the available context may be extended somewhat in this case.

<sup>14</sup> Only in Lagoudaki's earlier survey (2006), containing very diverse profiles, are specific occupations mentioned: translators use TMs most (39%), followed by project managers, reviewers and a small number of subtitlers (p. 12). For these last occupations no specific rates are provided.

<sup>15</sup> A fuzzy match is a segment in the TM that approximately or partly matches the segment in a new source text (Bowker & Barlow, 2004, pp. 72-73).

TMs	Corpora / corpus features
Risk of misaligned segments: manual verification of segments necessary during the alignment process	Segments are not extracted from their surrounding texts: the translator can still look at the text preceding or following the segments in question (no manual verification of segments needed)
Isolated sentences	Possibility of seeing the larger context while translating
The matching principle is based on the number of characters	Matches can be found based on semantics
Matches are displayed according to their ranks (automated decision-making process)	The translator is free to choose the best translation option
Mostly automatic copying and pasting of fuzzy matches or term matches	Translating from scratch is sometimes faster than editing an already existing translation
Sentence updating	Full text updating

**Table 7.** Comparison between TMs and corpora / corpus features

Alignment (linking source to target segments) is a crucial pre-processing step both in TMs and in (parallel) corpus querying. When using a TM, there is the risk “of ‘automatically’ retrieving misaligned segments” (Bowker & Barlow, 2008, p. 10). If this occurs in a bilingual concordancer (a tool which is aimed at finding all the occurrences of a search word or pattern in a corpus) the translator can still look at the text preceding or following the segment in question, because a concordancer “does not extract the segment from its surrounding text”, contrary to a TM (Bowker & Barlow, 2008, p. 10). Therefore, manual verification is not necessary prior to using a concordancer and this in turn saves time (Bowker & Barlow, 2008, p. 10).

As TMs store isolated sentences as translation units on the one hand, but context is an essential factor in translation on the other, implementing parallel corpora, including semi-automatic concordancing (cf. Bowker & Barlow, 2008), would allow translators to see the larger context. This is not possible when exclusively using a TM (Bowker, 2006, p. 179; Bowker & Barlow, 2008, p. 10; Frérot, 2010; Leblanc, 2013, p. 7; Tatu, 2011, p. 174) as a TM could be described as a repository of which all the sentences it contains are out of context (Bowker, 2006, p. 179). The segmented approach of TMs also changes the relationship translators have with the text, as it renders translation into a decontextualized activity. Often this approach influences “the quality of the final product in terms of syntagmatic cohesion and idiomaticity” (Leblanc, 2013, p. 7). Furthermore, this approach impedes creativity (Leblanc, 2013, p. 7).

With regard to matching, TM systems take into account similarity looking at the number of characters, leaving aside semantics (inflections, compounding, etc.). However, a parallel corpus and its concordancing tools will take this semantic dimension into account when proposing translation variants (Bowker & Barlow, 2008, p. 13; Tatu, 2011, p. 176). Further, contrary to TM software, parallel corpora and their concordancing tools display all possible matches, which leaves translators free to choose the best option(s). In this way they are not subject to a rather automatic decision-making process, as TMs display matches according to their ranks, and the highest-ranked match does not necessarily equal the best translation (Bowker & Barlow, 2008, p. 11; Tatu, 2011, p. 177). This automatic decision-making process makes translators lazy and passive, as TMs tend to be the only resource they consult or they just accept the TM suggestions without any questioning (Leblanc, 2013, p. 7).

Furthermore, many TM systems “automatically copy and paste fuzzy matches or term matches directly into the target text” (Bowker & Barlow, 2008, p. 11), but sometimes, typing a translation from scratch is faster than editing an automatically inserted segment from the TM. In this case the use of a bilingual concordance is more beneficial, as no text is automatically inserted in target documents, (Bowker & Barlow, 2008, p. 11). However, shadowing sessions of translators showed that some translators rather avoid translating from scratch, using “the ‘collage’ method, i.e., gathering translated sub-segments from the TM database and constructing sentences around those sub-segments” (Leblanc, 2013, p. 8). It is unclear how this affects translation quality.

Reusability is often claimed to be the top advantage of TM systems. However, updating sentences is likely to be more difficult and expensive than updating full texts, which is why a “more flexible, corpus-like approach” (Tatu, 2011, p. 178) should be adopted to the TM’s recycling procedure (Bowker & Barlow, 2008, p. 16). Some CAT tools, such as *MultiTrans* and *Logiterm*, have tried to incorporate a feature allowing to see a larger context by generating *bi-texts* (parallel corpora), which avoids the problem of decontextualized translation. This means that matches are linked to an entire document, instead of an isolated sentence (García, 2015, p. 72; Gow, 2003, p. 36). The aligned corpora generated in *MultiTrans* are fully accessible and can be searched while translating. Such a feature is mostly limited to parallel corpora at present. In addition, terminology extraction from different other types of corpora is also possible (cf. Frérot, 2010; cf. also Heylen & De Hertog, 2015, on Automatic Term Extraction or ATE). *SDL Trados* also includes corpus-like features, e.g., “project reference documents, *PerfectMatch* technology, and *AutoSuggest* dictionaries” (Mellinger, 2014, p. 22). Project reference documents are stored with the *SDL Trados* project files. The translator can open, view and search them if required while preparing for his translation. *PerfectMatch* technology allows for comparison between source files and existing bilingual (previously translated) documents instead of comparing source files to TMs (Mellinger, 2014, p. 22). *AutoSuggest* dictionaries are created in *SDL Trados* based on the analysis of TMs. Suggestions are made to the translator at the sub-sentential level which he can either use or discard (Mellinger, 2014, p. 23), with the aim of identifying “repetitive phrases or turns of phrase that can improve the translator’s ability to produce a translation that is similar in style and tone to previously translated material” (Mellinger, 2014, p. 24). To our knowledge the only CAT tool thus far incorporating the use of bilingual as well as monolingual corpora during translation is *MemoQ*, through its *Library™* technology.

## 5. Conclusions and further research

The suggestions and efforts established above with regard to the integration of corpora into CAT tools mostly concern aligned parallel and (to a lesser extent) comparable corpora (Bernardini and Castagnoli, 2008; García, 2015; Tatu, 2011).<sup>16</sup> However, Bowker (1998) showed that translation quality could also be improved by using specialized monolingual native-language (original) corpora, i.e., specialized corpora in one particular language written in the language of the native speaker. Corpora provide information other tools do not, particularly with regard to language use, which improves translation fluency (Loock, 2016, p. 2). With our survey data presented in Figure 4 (section 3.4) on freelance and in-house translators’ use of monolingual corpora in the source and target language for the language pairs English/Dutch and French/Dutch we aimed to provide a starting point, displaying the use of this type of corpus as a

<sup>16</sup> Cf. also Cristal project ([https://hal.archives-ouvertes.fr/hal-00986391/file/Josselin-LerayEtAl\\_Euralex2014.pdf](https://hal.archives-ouvertes.fr/hal-00986391/file/Josselin-LerayEtAl_Euralex2014.pdf)) with reference to the integration of contexts automatically retrieved from comparable corpora into CAT Tools (cf. also Picton, Planas & Josselin-Leray, 2017).

translation aid by our Belgian and Dutch respondents. To assess the effect of monolingual original corpora integrated into CAT tools a comparison of translations carried out with these monolingual original corpora and/or TMs is required, for instance with regard to coherence (context) and consistency. Such research into the integration of (monolingual) corpora and/or corpus-based features into the current world leading CAT tool (*SDL Trados*) could now also be extended as SDL has fairly recently acquired Donnelly Language Solutions, responsible for the *MultiTrans* tool (SDL plc, 2018). In addition, *SDL Trados* provides Application Programming Interfaces (APIs) to build applications which can help to optimize translation processes. An example of such an application which could be useful with regard to the integration of corpora into CAT tools is *Web Lookup* for instant web searches in *SDL Trados Studio* (“Taking translation to the next level: How customers use SDL APIs and applications to overcome the challenges of today’s translation industry”, SDL plc, 2017, pp. 13-14).

The main findings of our survey in terms of our hypotheses can be summarized as follows. Based on approximative indications of corpus use in the European and global recent surveys reviewed in section 2 we expected a higher frequency of corpus use among in-house translators compared to freelance translators. In that respect we did find higher adoption of the two other translation aids discussed in this paper, viz., CAT tools and TMs, among the in-house translators of our survey. The great majority of our respondents use a CAT tool, in line with the findings of the European and global surveys reviewed in section 2. But whereas especially in-house translators make most use of CAT tools in our survey, the frequencies are much more similar (and lower) for corpus use among *both* the freelance and in-house translators of our survey. Thus, for both translator profiles this may point to the need to integrate corpora better in their translation workflow systems and current tools, notably CAT tools. In addition, integrating corpus use in translation students’ curriculum, as illustrated well by Kübler et al. (2018), will probably lead to more corpus use among future translators.

From our survey parallel corpora still appear to be the most popular corpus type among those who indicated corpus use, as expected. Nevertheless, comparable and monolingual corpora may also provide benefits to be pursued in translation practice.

Concerning language pairs and corpus use we hypothesized that translators working with English make more frequent use of corpora as a translation aid, in view of the longer tradition of corpora and corpus linguistics for English compared to other languages. This hypothesis was clearly confirmed by our data.

We hope that the categorizations of our small-scale survey distinguished for the current paper may inspire further analysis based on larger datasets from greater respondent numbers, which might provide more conclusive evidence for the tentative conclusions which have been drawn here on the basis of our survey.

## 6. References

- Allard, M. G. P. (2012). *Managing terminology for translation using translation environment tools: Towards a definition of best practices* (unpublished doctoral dissertation). University of Ottawa, Canada.
- Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223-243.
- Bernardini, S. & Castagnoli, S. (2008). Corpora for translator education and translation practice. In E. Yuste Rodrigo (Ed.), *Topics in language resources for translation and localisation* (pp. 39-56). Amsterdam: Benjamins.
- Blancafort, H. & Gornostay, T. (2010). Calling professionals: Help us to understand your needs. Retrieved from [www.researchgate.net/profile/Tatiana\\_Gornostay/publication/236626809\\_Calling\\_Professionals\\_Help\\_Us\\_to\\_Understand\\_Your\\_Needs/links/02e7e51882f0b3d6f5000000/Calling-Professionals-Help-Us-to-Understand-Your-Needs.pdf](http://www.researchgate.net/profile/Tatiana_Gornostay/publication/236626809_Calling_Professionals_Help_Us_to_Understand_Your_Needs/links/02e7e51882f0b3d6f5000000/Calling-Professionals-Help-Us-to-Understand-Your-Needs.pdf)

- Bowker, L. (1998). Using specialized monolingual native-language corpora as a translation resource: A pilot study. *Meta*, 43(4), 631-651.
- Bowker, L. (1999). Exploring the potential of corpora for raising language awareness in student translators. *Language awareness*, 8(3-4), 160-173.
- Bowker, L. (2004). Corpus resources for translators: Academic luxury or professional necessity? *TradTerm*, 10, 213-247.
- Bowker, L. (2005). Productivity vs quality. A pilot study on the impact of translation memory systems. *Localisation Focus*, 4, 13-20.
- Bowker, L. (2006). Translation memory and “text”. In L. Bowker (Ed.), *Lexicography, terminology and translation* (pp. 175-187). University of Ottawa Press.
- Bowker, L. (2015). Terminology and translation. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology, Vol. 1*. (pp. 304-323). Amsterdam: Benjamins.
- Bowker, L. & Barlow, M. (2004). Bilingual concordancers and translation memories: A comparative evaluation. In *Proceedings of the second international workshop on language resources for translation work, research and training*. Geneva, Switzerland, 28 August 2004 (pp. 70-79). Stroudsburg, PA: Association for Computational Linguistics.
- Bowker, L. & Barlow, M. (2008). A comparative evaluation of bilingual concordances and translation memory systems. In E. Yuste Rodrigo (Ed.), *Topics in language resources for translation and localisation* (pp. 1-22). Amsterdam: Benjamins.
- CBTI-BKVT (2018). *Market survey report*. Retrieved from [https://aticom.de/wp-content/uploads/2018/09/CBTI-BKVT\\_Market-Survey-Report-2018.pdf](https://aticom.de/wp-content/uploads/2018/09/CBTI-BKVT_Market-Survey-Report-2018.pdf)
- Dillon, S. & Fraser, J. (2006). Translators and TM: An investigation of translators’ perceptions of translation memory adoption. *Machine Translation*, 20(2), 67-79.
- Fantinuoli, C. (2016). Revisiting corpus creation and analysis tools for translation tasks. *Cadernos de Tradução*, 36(1), 62-87.
- Fernández-Parra, M. (2010). The workflow of computer-assisted translation tools in specialised translation. In C. Heine & J. Engberg (Eds.), *Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009*. Retrieved from <http://bcom.au.dk/fileadmin/www.asb.dk/isek/fernandez-parra.pdf>
- Frankenberg-Garcia, A. (2015). Training translators to use corpora hands-on: Challenges and reactions by a group of thirteen students at a UK university. *Corpora*, 10(3), 351-380.
- Frérot, C. (2010). Outils d’aide à la traduction : pour une intégration des corpus et des outils d’analyse de corpus et dans l’enseignement de la traduction et la formation des traducteurs. *Les Cahiers du GEPE 2*. Retrieved from [www.cahiersdugepe.fr/index.php?id=1164](http://www.cahiersdugepe.fr/index.php?id=1164)
- Fulford, H. & Granel-Zafra, J. (2005). Translation and technology: A study of UK freelance translators. *The Journal of Specialised Translation*, 4(1), 2-17.
- Gallego-Hernández, D. (2015). The use of corpora as translation resources: A study based on a survey of Spanish professional translators. *Perspectives*, 23(3), 375-391.
- García, I. (2006). Translators on translation memories: A blessing or a curse? In A. Pym, A. Perekrestenko & B. Starink (Eds.), *Translation technology and its teaching (with much mention of localization)* (pp. 97-105). Tarragona: Universitat Rovira i Virgili.
- García, I. (2015). Computer-aided translation systems. In C. Sin-wai (Ed.), *The Routledge encyclopedia of translation technology* (pp. 68-97). New York: Routledge.
- Gough, J. (2011). An empirical study of professional translators’ attitudes, use and awareness of Web 2.0 technologies, and implications for the adoption of emerging technologies and trends. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, 10, 195-225.
- Gow, F. (2003). *Metrics for evaluating translation memory software (unpublished master thesis)*. School of Translation and Interpretation, University of Ottawa, Canada.
- Heylen, K. & De Hertog, D. (2015). Automatic term extraction. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology, Vol. 1* (pp. 203-221). Amsterdam: Benjamins.
- Johansson, S. (2007). *Seeing through multilingual corpora: On the use of corpora in contrastive studies*. Studies in corpus linguistics (Vol. 26). Amsterdam: Benjamins.
- Kübler, N. (2003). Corpora and LSP translation. In F. Zanettin, S. Bernardini & D. Stewart (Eds.), *Corpora in translator education* (pp. 25-42). Manchester: St. Jerome.
- Kübler, N., Mestivier, A. & Pecman, M. (2018). Teaching specialized translation through corpus linguistics: Translation quality assessment and methodology evaluation and enhancement by experimental approach. *Meta*, 63(3), 807-825.
- Lagoudaki, E. (2006). Translation memories survey 2006: Users’ perceptions around TM use. In *Proceedings of the ASLIB International Conference Translating & the Computer 28* (pp. 1-29). London: Aslib.

- LeBlanc, M. (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *Translation & Interpreting*, 5(2), 1-13.
- Lock, R. (2016). L'utilisation des corpus électroniques chez le traducteur professionnel: quand? comment? pour quoi faire? *Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 27. Retrieved from <https://ilcea.revues.org/3835>
- MeLLANGE Multilingual eLearning in LANGuage Engineering. (2006). *Corpora & e-learning questionnaire results summary (2006)*. Retrieved from <http://mellange.eila.univ-paris-diderot.fr/Mellange-Results-1.pdf>
- Mellinger, C. D. (2014). *Computer-assisted translation: An empirical investigation of cognitive effort* (unpublished doctoral dissertation). Kent State University, Ohio, USA.
- Metatexi.com. (n.d). Why translate with a CAT tool? Accessed August 25, 2019, <http://www.metatexi.com/cat.htm>
- Moorkens, J. (2012). *Measuring consistency in translation memories: A mixed-methods case study* (unpublished doctoral dissertation). Dublin City University.
- Popiolek, M. (2015). Terminology management within a translation quality assurance process. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology, Vol. 1* (pp. 341-359). Amsterdam: Benjamins.
- Picton, A., Fontanet, M., Maradan, M. & Pulitano, D. (2015, May). *Corpora in translation: Addressing the gap between the scholar's and the translator's point of view*. Presentation at the International Conference on Corpus Use and Learning to Translate (CULT), Alicante, Spain.
- Picton, A., Planas, E. & Josselin-Leray, A. (2017). Monitoring the use of newly integrated resources into CAT tools: A prototype. In A. Corpas & I. Durán-Munoz (Eds.), *Trends in e-tools and resources for translators and interpreters, Approaches to translation studies* (pp. 109-136). Leiden: Brill Editions.
- SDL plc. (2018). Completion of acquisition of Donnelley language solutions. Retrieved from <https://www.sdl.com/about/news-media/press/2018/completion-of-acquisition-of-donnelley-language-solutions.html>
- SDL plc. (2017). *Taking translation to the next level: How customers use SDL APIs and applications to overcome the challenges of today's translation industry*. Retrieved from [www.sdltrados.com/resources/ebooks/](http://www.sdltrados.com/resources/ebooks/)
- Steurs, F., De Wachter, K. & De Malsche, E. (2015). Terminology tools. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology, Vol. 1* (pp. 222-249). Amsterdam: Benjamins.
- Steurs, F. & van der Lek-Ciudin, I. (2016). Smart computer aided translation environment report on human terminology extraction – Deliverable D3.1. KU Leuven, Campus Sint-Andries Antwerp, Antwerp.
- Tatu, O. (2011). Using CAT systems and/or corpora in the translation profession. *Revue roumaine de linguistique*, 56(2), 167-180.
- Vandeghinste, V., Vanallemeersch, T., Augustinus, L., Bulté, B., Van Eynde, F., Pelemans, J., . . . Luyten, K. (2019). Improving the translation environment for professional translators. *Informatics*, 6(2). <https://doi.org/10.3390/informatics6020024>
- Van den Bergh, J., Geurts, E., Degraen, D., Haesen, M., van der Lek-Ciudin, I. & Coninx, K. (2015). Recommendations for translation environments to improve translators' workflows. In J. Esteves-Ferreira, J. M. Macan, R. Mitkov & O.-M. Stefanov (Eds.), *Proceedings of the 37th Conference Translating and the Computer* (pp. 106-119). Geneva: Tradulex.
- Wilkinson, M. (2010). Quick corpora compiling using web as corpus. *Translation Journal*, 14(3). Retrieved from <http://translationjournal.net/journal/53corpus.htm>
- Zaretskaya, A., Corpas Pastor, G. & Seghiri, M. (2015). Translators' requirements for translation technologies: A user survey. In G. Corpas Pastor, M. Seghiri Domínguez, R. Gutiérrez Florido & M. Urbano Mendaña (Eds.), *Proceedings of the AIET17 Conference. New Horizons in Translation and Interpreting Studies* (pp. 247-254). Geneva: Tradulex.
- Zaretskaya, A., Corpas Pastor, G., & Seghiri, M. (2018). User perspective on translation tools: Findings of a user survey. In G. Corpas Pastor & I. Durán-Muñoz (Eds.), *Trends in e-tools and resources for translators and interpreters* (pp. 37-56). Leiden: Brill/Rodopi.
- Zielinski, D., & Ramirez, Y. (2005). *Research meets practice: t-survey 2005*. Retrieved from <http://mt-archive.info/Aslib-2005-Zielinski.pdf>

## Appendix 1 : Complete Survey

### Survey: The use of CAT tools and corpora by professional translators

#### Introduction [.....]

#### Definitions

What follows is a clarification of several terms used in the survey, in order to avoid ambiguity.

- CAT tool: A CAT (computer-assisted translation) tool supports the translation process through software. Usually a CAT tool consists of several components, such as a translation memory (TM) and a terminology management system. A translation memory is a database which stores source segments and their translated target segments. Translation memories can be created from scratch in the CAT tool itself or translation memories can be imported in your CAT tool to retrieve segments immediately. Using translation memories makes the translation process faster and more efficient. A few examples of CAT tools are SDL Trados, MemoQ and Wordfast.
- Terminology management: Terminology management in this survey refers to the collection, storage and retrieval of terms. Terminology management can be performed using standard tools such as Microsoft Word or Excel. Specific terminology management systems are also available, which can be used stand-alone or can be integrated in a CAT tool. An example of a terminology management system is SDL MultiTerm, which can be integrated in SDL Trados.
- Termbase: A database containing terminology.
- Corpus: A corpus in this survey refers to a collection of written texts which is used for searching equivalent terms while translating. A corpus which contains original texts in one particular language is called a mono-lingual corpus. A corpus including collections of texts with similar content in different languages or language varieties (e.g., British English, American English, Australian English) is called a comparable corpus. A comparable corpus does not consist of original texts along with their corresponding translations (cp. parallel corpus). Therefore, the (linguistic) content is not influenced by translation processes. A parallel corpus consists of original texts in one particular language, along with their translations in a different language. Examples of parallel corpora on the web are Linguee and Reverso Context, which are ready-made corpora. However, translators can also compile their own corpora.

#### Part I: General Information

##### Q1: What is your age group?

- 18-30 (A1)      ◦ 31-40 (A2)      ◦ 41-50 (A3)      ◦ 51-60 (A4)      ◦ 60+ (A5)

##### Q2: What are your main qualifications?

- Professional bachelor degree in Translation
- Master degree in Translation
- Master degree in Interpreting
- Master degree in Languages (or equivalent)
- Specialized courses, seminars, etc.
- Other, please specify:

##### Q3: Which of the following best describes your job role?

- In-house translator
- Freelance translator

##### Q4: How many years of translation experience do you have?

- Less than 1 year                      ◦ 1 - 5 years                      ◦ 6 - 10 years
- 11 - 15 years                      ◦ 16 - 20 years                      ◦ 20+ years

**Q5: Which language combination(s) do you work with?**

From ...	Into ...
◦ French	◦ Dutch (SQ001)
◦ English	◦ Dutch (SQ002)
◦ German	◦ Dutch (SQ003)
◦ Spanish	◦ Dutch (SQ004)
◦ Dutch	◦ French (SQ005)
◦ Dutch	◦ English (SQ006)
◦ Dutch	◦ German (SQ007)
◦ Dutch	◦ Spanish (SQ008)
◦ Other(s), please specify:	

**Q6: Which text domain(s) (if any) do you specialize in?**

- I do not specialize in any domain.
- Administration (SQ001)
- Engineering (SQ002)
- Health (SQ003)
- Finance (SQ004)
- Information Technologies (SQ005)
- Marketing (SQ006)
- Law (SQ007)
- Tourism (SQ008)
- Other(s), please specify:

**Part II: CAT tools**General use**Q7: Do you use (a) CAT tool(s)?**

- Yes    ◦ No

(if the answer is *Yes*, continue to Q8; if the answer is *No*, continue to Q12)

**Q8: Which CAT tool(s) do you use?**

- Across (A1)
- Alchemy CATALYST (A2)
- DéjàVu (A3)
- Lionbridge Translation Workspace (A4)
- MateCat (A5)
- MemoQ (A6)
- MemSource (A7)
- MultiTrans (A8)
- Similis (A9)
- Swordfish (A10)
- SDL Trados (A11)
- Transit (A12)
- OmegaT (A13)
- Wordbee (A14)
- Wordfast (A15)
- Other(s), please specify:

**Q9: Have you received any training on CAT tools?**

- Yes    ◦ No

(if the answer is *Yes*, continue to Q10; if the answer is *No*, continue to Q11)

**Q10: What kind of training on CAT tools did you receive?**

- Training course(s) at school (SQ001)
- In-house training course(s) (SQ002)
- External training through workshops, seminars, webinars, etc. (SQ003)
- I am self-taught in the use of CAT tools and did not receive any training. (SQ004)
- Other, please specify: (SQ005)

**Q11: Which of the following tasks do you perform with CAT tools?**

- Translation (SQ001)
- Translation consistency check (SQ002)
- Translation completeness check (SQ003)
- Import of translation memories (SQ004)
- Text analysis (e.g., for invoicing) (SQ005)
- Terminology management (SQ006)
- Terminology quality assessment (SQ007)
- Other(s), please specify: (SQ008)

**Q12: Why do you not use CAT tools?**

- They are too expensive.
- They are too complicated.
- I feel I do not need them.
- Other, please specify:

**Q13: Are you planning to use CAT tools in the future?**

- Yes
- No

Terminology management**Q14: Do you engage in terminology management?**

- Yes (A1)
- No (A2)

(if the answer is *No*, continue to Part III: Corpora)

**Q15: Which resources do you use to manage terminology?**

- Microsoft Word (SQ001)
- Microsoft Excel (SQ002)
- Terminology management system (stand-alone or integrated in a CAT tool) (SQ003)
- Other(s), please specify:

**Q16: Which terminology management system(s) do you use?**

(only answer if *Terminology management system* was selected in Q15)

- crossTerm (Across) (SQ001)
- qTerm (MemoQ)
- SDL MultiTerm (Trados)
- Star Termstar (Transit)
- Other(s), please specify:

**Q17: How do you mostly obtain terminology?**

(Only answer if *Yes* was selected in Q14)

- I compile my own termbases/terminology lists based on source texts and their translations for future use. (A1)
- I obtain terminology from clients which I use when doing translations for them.
- I obtain terminology from other resources (the Internet, literature, etc.) which I then use to fill out a term-base/terminology list.
- Other, please specify:

### Part III: Corpora

**Q18: Do you use corpora in translation?**

- Yes (A1)
- No (A2)

(if the answer is *Yes*, continue to Q19; if the answer is *No*, continue to Q22)

**Q19: What kind of corpora do you use? Select all that apply.**

- Monolingual corpora in the source language
- Monolingual corpora in the target language
- Comparable corpora
- Parallel corpora

**Q20: Do you build your own corpora for translation?**

- Yes (A1)
- No (A2)

(if the answer is *Yes*, continue to *Any additional remarks*, if the answer is *No*, continue to Q21)

**Q21: How do you build your corpora, e.g., which tools do you use for corpus building?**

**Q22: Would you be interested in building your own corpora for translation?**

- No (A1)
- Yes, please provide your e-mail address: \_\_\_\_\_

**Q23: Why do you not use corpora?**

- I am not familiar with them. (A1)
- They are too time-consuming to compile. (A2)
- I feel I do not need them. (A3)
- Other, please specify:

**Q24: Are you planning to use corpora in the future?**

- No (A1)
- Yes. Please provide your e-mail address: \_\_\_\_\_

### Part IV: Miscellaneous

**Q25: Any additional remarks:**

---

---

**Q26: Please provide your e-mail address if you are interested in receiving the results of this survey.**

---

---



 Heidi Verplaetse

KU Leuven, Humanities and Social Sciences  
Research Group QLVL – Research Unit Linguistics  
Campus Sint-Andries  
Sint-Andriesstraat 2  
2000 Antwerp  
Belgium

[Heidi.Verplaetse@kuleuven.be](mailto:Heidi.Verplaetse@kuleuven.be)

**Biography:** Heidi Verplaetse currently lectures English at KU Leuven, Campus Antwerpen in the domains of translation and writing (medical/scientific, business, journalistic) and English external communication. She also worked in the domain of knowledge management based on automatic retrieval of semantic relations and anaphora. Her PhD in English linguistics (Ghent University, 2008) was situated in the domain of modality. Her current research interests are situated in the domains of specialized (medical) translation, translation quality and the use of CAT tools, trans-editing of journalistic texts, translation quality and processes in students, and the expression of epistemic modality in historical scientific writing.

#### An Lambrechts

KU Leuven, Humanities and Social Sciences  
Research Group QLVL – Research Unit Linguistics  
Campus Sint-Andries  
Sint-Andriesstraat 2  
2000 Antwerp  
Belgium

[An.Lambrechts@workmail.com](mailto:An.Lambrechts@workmail.com)

**Biography:** An Lambrechts obtained her MA degree in Translation at KU Leuven, Campus Antwerp. She has worked as a professional in-house translator for ten years, specializing in medical, legal and technical translation, before embarking on a research project in translation and linguistics at KU Leuven in September 2016. The focus of her research is the impact of CAT tools and corpora on the quality and efficiency of specialized (medical/legal) translations from English into Dutch. Her research interests include specialized translation, phraseology, corpus linguistics and translation technologies. She currently also works as a technical writer.



This work is licensed under a Creative Commons Attribution 4.0 International License.