Universidad de Valladolid

FACULTAD de FILOSOFÍA Y LETRAS

DEPARTAMENTO de FILOLOGÍA INGLESA

Grado en Estudios Ingleses

TRABAJO DE FIN DE GRADO

COMPARATIVE ANALYSIS OF

COMPUTER-ASSISTED TRANSLATION TOOLS

Sergio Tarilonte Pérez

Tutora: Belén López Arroyo

2018/2019

ABSTRACT

Nowadays, information technologies are growing rapidly to the point of supplying the real world in many ways. In the world of translation this is also happening with the constant irruption of new programs and web resources that allow an increasingly perfect automatic translation of texts. However, these tools are still far from perfect and, until they are, the best resource that can be used by a translator is computer assisted translation (CAT) tools. This paper tries to discern what each of the most important CAT tools offers, among which any translator can harbor doubts which is more convenient, and to establish a comparison between them. The results will give the reader a vision of both, which tool is more convenient to use, and of what it is that it provides the user when developing a translation.

Key words: Translation Studies, Computer-Assisted Translation, Translation memories, Program Comparison, Machine Translation, Parallel Corpora.

En la actualidad, las tecnologías de la información están creciendo rápidamente hasta el punto de llegar a suplir al mundo real en muchos aspectos. En el mundo de la traducción esto también está sucediendo con la irrupción constante de nuevos programas y recursos web que permiten una traducción de textos automática cada vez más perfecta. Sin embargo, estas herramientas aún distan de ser perfectas y, hasta que lo sean, el mejor recurso que puede emplear un traductor son las herramientas de traducción asistida por ordenador (TAO). Este trabajo trata de discernir qué ofrece cada una de las más importantes herramientas TAO, entre las que cualquier traductor puede albergar dudas de cual es más conveniente, y establecer una comparativa entre estas. Los resultados servirán para dar al lector una visión tanto de, qué herramienta es más conveniente usar, como de qué es lo que le aporta al usuario a la hora de elaborar una traducción.

Palabras clave: Estudios de Traducción, Traducción Asistida por Ordenador, Memorias de Traducción, Comparación de programas, Traducción Automática, Corpus Paralelos.

INDEX

1. INTRODUCTION

Thirty years ago, the first commercial translation memory tools appeared in everyday life (Lagoudaki, 2006) (See Section 2 below). Nowadays, these tools are still improving constantly with new elements and techniques, like the capability of using online translators, that make the task of the translators, both professional and non-professional, easier. This constant evolution makes this technology an interesting field of research.

This paper tries to establish an analysis of the features of some of the main Computer Assisted Translation tools, that use translation memories, showing how those tools work and what they can offer to the target users, aiming to create a critical point of view in the reader of the advantages and disadvantages of these programs beyond its price or popularity.

Since there are too many Computer Assisted Translation tools (CAT tools), this paper will focus on some of the most used ones to analyze them properly, since the size of this paper is not large enough to analyze all programs in depth.

To check which ones are better in order to elaborate the analysis, this paper uses the following chart (Figure 1) elaborated in 2013 by Jared Tabor in a survey made to over three thousand translators about their CAT tools preferences.
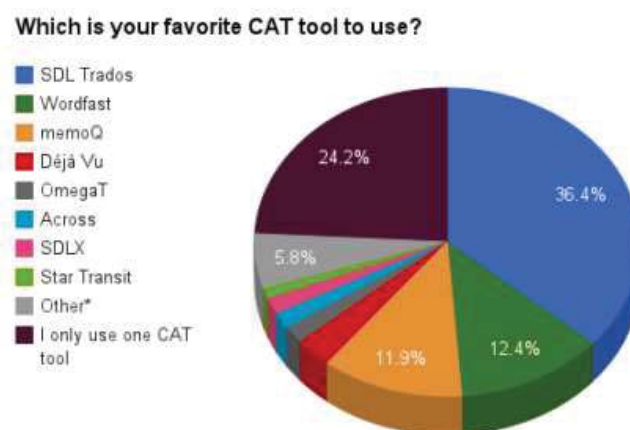

Figure 1: Favorite CAT tools

This pie chart stands out among the others made by Tabor (2013) because of an important aspect; the survey respondents have used more than one CAT software, so their opinions are not biased by some facts like the price or the habit. The price of these tools is usually expensive, so the users don't want to expend any extra money on a new one, or once they get used to a software, they won't likely switch to a different one unless they find a very good reason.

As it can be seen in Figure 1 the five most preferred tools are Trados, Wordfast, MemoQ, Déjà Vu and OmegaT which will be the ones selected for this paper as they offer at least a trial version to work with.

In order to understand this work as clearly as possible, this paper is divided into four main sections.

The first section of this paper, Fundamentals of Machine Translation, tries to explain the basic concepts necessary to understand this topic. This section is one of the main ones of this work because it briefly explains the complex or difficult to understand concepts that will be discussed in the paper.

Shorter than the previous one, the second section, Methodology, deals with the description of the different processes needed to begin the comparative analysis in the following section along with a brief explanation of how the subsequent analysis will be carried out.

After that explanation, the third section, and the most important segment of this paper, Analysis and Evaluation, tries to perform an individual study of the five Computer Aided Translation tools preferred by the translators along with a later joint comparison of those tools and a final interpretation of all the collected data.

Finally, the fourth and last section of this paper, Conclusion, summarizes all the work carried out and extracts the final results of the research carried out.

2. FUNDAMENTALS OF MACHINE TRANSLATION

To understand this paper is important to know some basic notions of translation memories and Computer Assisted Translation tools.

2.1. What is a corpus

In order to begin with this brief explanation of the basic concepts necessary to understand this paper, there is a key concept that, as will be seen below, will be repeated throughout the document. This is the concept of corpus (corpora in plural).

According to many dictionaries, as Oxford, Cambridge or Collins, a corpus, in the field we want to address, is "a collection of material, spoken or written, in machine-readable form, assembled for the purpose of finding out how language is used".

The different kinds of corpora had been classified in many ways by various authors over time. In figure 2 it can be seen an attempt of classification elaborated by Ornia (1996). In her paper, she studies the lack of agreement in the classification of corpora, trying to elaborate her own classification based on the proposals of different authors, as shown below. Her classification divides different corpora in branches resulting in two different branch tree diagrams.

The first diagram comprehends mostly qualities common to all corpora such as if their source is in oral or writing format, if they are or not representative of a language (reference corpus), if they are fixed to a period of time or not (synchronic or diachronic corpora), if the corpora could be actualized with new information (only in the case of diachronic corpora, as they are not fixed to a period of time), if their source are full texts, fragments or a combination of both, if those texts are from published or unpublished sources and if their purpose is general or specialized to a topic in particular.

The second diagram continues from the division among general and specialized corpora dividing the corpora regarding if they have one or more languages (monolingual and multilingual corpora). At this point, she suggests a division between parallel and comparable corpora, but, as she exposes, the parallel corpora must be the original text along with their respective translations whereas the comparable corpora could be comprised of original and translated texts in one or more languages that do not fulfill the

requirement of the parallel corpora, as stated in the final branch of the second classification of figure 2.

Once this division is established, Ornia continues explaining that parallel corpora can be classified between unidirectional, if all translations go from one A language to another (b, c, d) and multidirectional languages, if the translations go indistinctly from one language to another, and she also concludes her classification explaining that parallel corpora can be aligned or unaligned, but in turn leaves room for future improvements in the classification such as the mother languages, the translation methods, the translator level, if they are open….

In order to complete and give a little more depth to this concept and its classification it is advisable to finish this section showing some of the most remarkable examples of corpora, among which we could highlight the ones shown by Berglund Prytz (2012).

The first remarkable type of corpora, balanced corpora, "try to represent a particular type of language over a specific span of time. In doing so they seek to be balanced and representative within a particular sampling frame" (McEnery and Hardie, 2012). The main exponent of this type of corpus is the British National Corpus (BNC) that includes thousands of sentences and words in British English extracted and categorized according to its source.

The monitor corpora, tries on the other hand, as its name suggests, to monitor the changes in a specific language during different periods of time. Some of the maximum exponents of this type of corpus are the Corpus of Contemporary American English (COCA), which includes fragments of texts that go from 1990 to the present days, or the Bank of English (BoE) that includes different categories of modern texts and recordings mainly extracted from newspapers and media.

Another kind of corpora that Berglund Prytz suggests is the diachronic corpora collects comparable texts from various periods of a language. A good example of this type of corpus is the Helsinki Corpus of English Texts that includes a huge amount of Old, Middle and Early Modern English texts. As this example does not go beyond the Early Modern English it would not be correct to label it as a monitor corpus and would be more in the category of non-monitored diachronic corpus exposed by Ornia (1996).
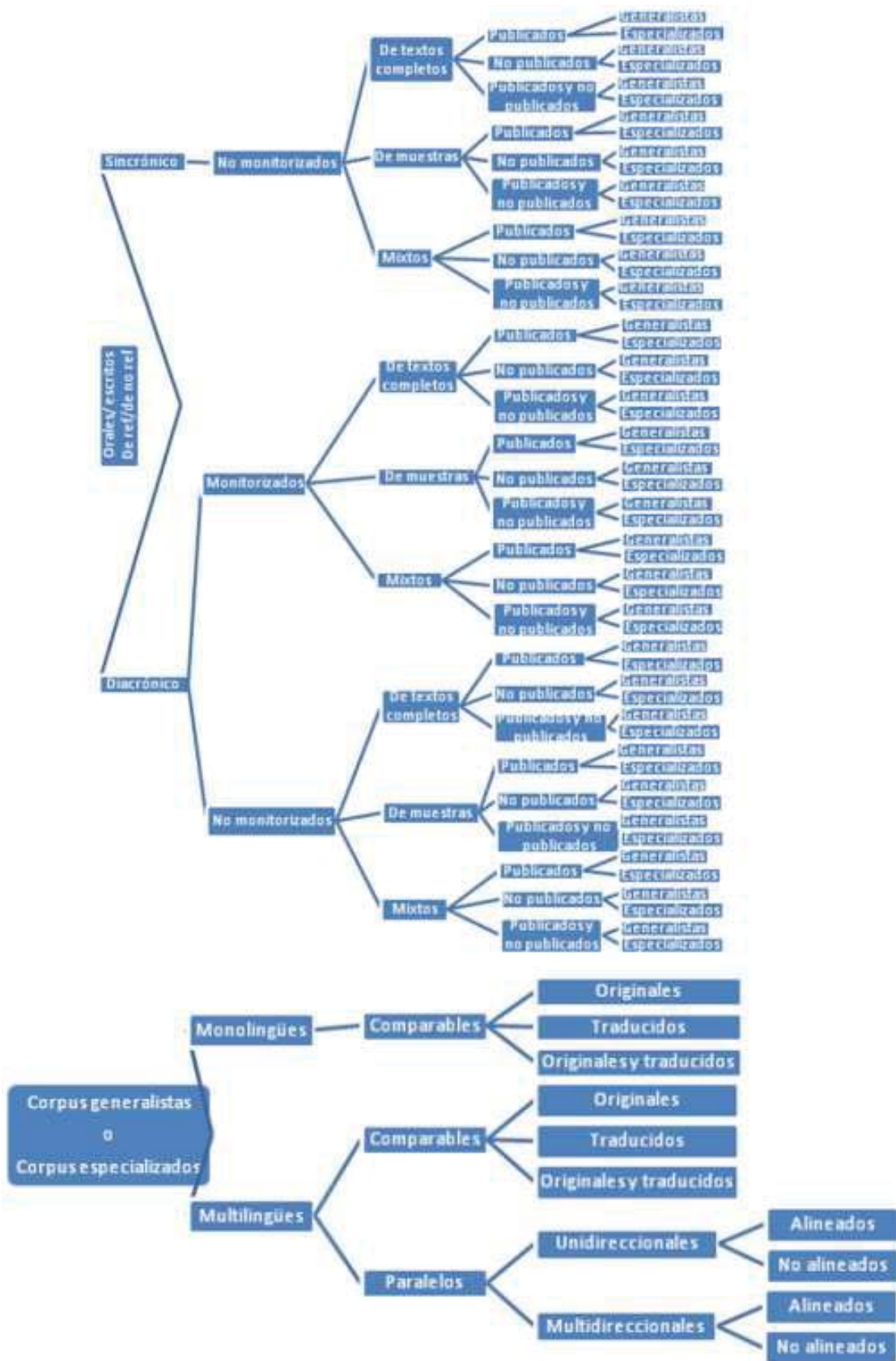
Figure 2: Corpora classification (Ornia,1996)

Other good example of corpora can be found on the International Corpus of English (ICE) which includes texts of various varieties of English throughout the world such as Irish English, Indian English or African English. As it include texts in different languages, or varieties of a language, of the same genres in the same domains and it uses the same sampling method for all the texts it fits perfectly in the multilingual comparable category exposed by Ornia (1996)and the definition of a comparable corpus exposed by McEnery and Hardie (2014).

On the other hand, as an example of parallel corpora could be highlighted the open source parallel corpus (OPUS) or the corpus prepared for the web-based translator Linguee. These kinds of corpora are composed only by original texts and their respective translations. The corpus developed for the study made in this paper belongs mainly to this category.

To conclude another type of remarkable corpus that have been used in one of the articles mentioned in this paper, is the TURICOR used by Corpas Pastor and Seghiri (2010) (section 2.4) that fits mainly as a specialized corpus as it focus on a specific kind of text (tourism in this case). As the corpus used in this paper is focused also on a specific kind of text (instruction manuals), apart from multilingual parallel corpus, it could also be defined as a specific corpus.

2.2. What is Machine Translation?

Once we have briefly explained the concept of corpus and some of the many examples and different types that can be found, it is important to focus on what is machine translation so that later we can focus on how it is classified in order to see how it has evolved and introduced the corpora in the most recent translation systems.

According to SDL, developer company of the first worldwide CAT tool, Trados, Machine translation, also known as automated or instant translation, is "The translation of text by a computer, with no human involvement.". But, in the words of Chiew Kin Quah (2006), Machine Translation is even more than this. She exposes that machine translation is also an important technology in the scientific, commercial and socio-political spheres that has become a bridge between languages with the emergence of the Internet.

For Quah (2006), Machine Translation is an enterprise that combines a large number of fields of study such as linguistics, computer science, lexicography, language engineering… and it is a great alternative to the human translation in many situations as it is more cost-effective. It is based on the theory that languages can be mathematically coded completely.

2.3 Classification of translation systems

To understand how Machine Translation works many authors divide this kind of translation in three different systems: rule-based, statistical, and neural.

The oldest of these systems, the rule-based system, combine grammar rules with dictionaries in order to translate common words. This approach contains a combination of syntactic, morphologic and semantic rules to analyze a Source Language Text (SLT) and, later on, synthetizing it into a Target Language Text (TLT) (Carl and Way 2003). In other words, this translation system analyzes the meaning representation of the SLT generating an equivalent to that meaning in the TLT.

The statistical machine translation system, as its name suggests, uses the probability to combine the information extracted from previous translations in order to perform future translations. According to SDL, this system makes better translation in terms of fluency, but, on the other hand, its translations tends to be less consistent.

To perform statistical-based approach translations, the SLT must be segmented into short strings of phrases or even words, then those segments are then compared with an existing bilingual corpus made of original texts and their translations, and then a statistical method based on the Bayes' theorem is employed on the aligned bilingual corpus to achieve the new target language segments (Carl and Way 2003).

This approach is based on the hypothesis that a source language sentence could have many different translations, each one with a different probability of being correct. Bayes states that this probability can be calculated through the following formula:

$$P(T|S) = \frac{P(T) \times P(T|S)}{P(S)}$$

Figure 2: Bayes' theorem

This approach, though it is better than the rule-based one, still have many problems, as has been stated previously. Its consistency depends on a big bilingual corpus made of strings of a small number of words without contextual information, or n-grams, to perform good translations. This lack of contextual information could also perform several mistakes as the tool cannot understand only with statistics the context in which some expressions are used.

To solve this issue most of the current translation tools combine these two different approaches in order to compensate the shortages of both while the third approach is being developed.

This third approach improves the statistical corpus-based machine translation employing a large neural network based on the processes made by the brain. Unlike the statistical machine translation, which divides its process into two separate steps, alignment and phrase extraction, neural machine translation takes all the process into a single process where all its components are integrated. Due to that, the neural machine translation does not only take into account the previous words of each segment in order to translate the following words but taking all the segment words into account (Forcada 2017). Google Translate is a good example of this kind of machine translation.

2.3. What is CAT tool?

Even though the main target of machine translation is to achieve the Fully Automatic High-Quality Translation (FAHQT), this is still far from being achieved and human translators will be still necessary. Hutchins and Somers (1992) discuss that, between Human Translation and FAHQT, there are two categories that employ both human and machine translation: Human-Aided Machine Translation (HAMT), where the translator aids during the translation computerized processes, and Machine-Aided Human Translation (MAHT), where the computer is the one that aids the professional translator. As they think that sometimes it could be difficult to separate these two categories, they employ the term Computer-Aided Translation, or Computer-Assisted Translation, (CAT) to comprehend both categories.
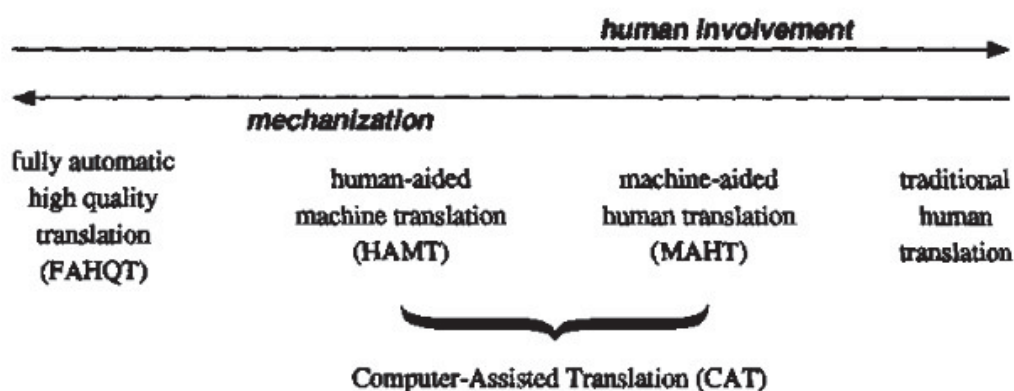
Figure 3: Classification of translation types (Hutchins and Somers, 1992)

According to Fernández-Rodríguez (2010, p. 205) "a CAT system integrates several functionalities represented by resources that allow, as a minimum, the terminological management, the storage and reuse of previously translated segments (in translation memories), and the alignment of texts.", therefore, these two categories had in common the use of translation memories.

At this point, it bears asking what is a translation memory. A translation memory is a database that automatically collects fragments of previous translations in order to translate strings of words, which match, totally or partially, to the already translated segments (Zanettin 2002). These databases are some kind of very specialized parallel corpora that extremely help in the task of translating future text related to the same topic however, they become less useful upon translating less related topic and its way of translating sentences could become very repetitive to the reader. One of the tasks of the translators is to avoid these shortages by feeding the translation memory and post editing the final text.

2.4. Representativity of corpora

Finally, as the final purpose of this paper is to analyze some of the different CAT tools that are on the market and, for this, one of the parts of the comparison consists of an empirical analysis based on a reference corpus, it is necessary analyze if this corpus is relevant enough to carry out the analysis, as seen in the next section of this paper. Therefore, another important aspect in order to understand this paper is the representativity of corpora.

When compiling a new corpus, it is very important to know when the corpus is wide enough to be representative. This point has been a debate subject among several authors. For example, Haan (1992) exposed that the minimum number of words on a corpus to be representative does not exceed the twenty thousand words, or Biber (1995), which reduced that minimum to one thousand words in exactly ten different texts.

In order to establish the minimum representative number of texts Biber (1995), Heaps (1971), Leech (1991) and many other authors established their approaches using formulas based on Zipf's law, which divide the components of the text in tokens, the number of all the words in the text, types, the number of words excluding repetitions, and the frequency of words extracted from divide types into tokens. His law stated that the number of occurrences of a word f(n) is inversely proportional to its or number on the frequency list (k), being the last element in *Figure 5* (s) the exponent value that characterizes the classification (Seghiri, 2014).

$$f(k; s, N) = \frac{1/k^s}{\sum_{n=1}^{N} 1/n^s}$$

Figure 5: Zipf's law

On the other hand, Corpas Pastor and Seghiri (2010), keeping in mind the studies from their predecessors, looked for a way to calculate mathematically the representativity and, therefore, suggest the use of Lexitrad's ReCor as a useful solution to determine the minimum size of a corpus regardless of its the language or genre, and establishing its minimum threshold of representativeness through an analysis of the corpus' lexical density using the following N-Cor algorithm.

$$C_n = d_1 + d_2 + d_3 + ... + d_n$$

Figure 6: N-Cor algorithm

3. METHODOLOGY AND CORPUS

Once the bases of this paper have been established, whose main objective is to establish a comparative analysis of some of the most used CAT tools, it is also important to establish the processes necessary to carry out the analysis. This section tries to explain these processes by first, checking the relativity of the corpus, given the reasons cited in the previous section, and later, the explanation of the methodology to be used in the project.

3.1. Representativity of corpus

As showed in the previous section, CAT tools need a related translation memory to work properly. In order to carry out the subsequent analysis of CAT tools, it is necessary to feed those tools with a translation memory broad enough to perform their translations.

To fulfill the task of this paper there has been selected a group of nine different instruction manuals of different devices from around 2012 and 2018 related with the topic of television systems (Specifically, the sources used are: the instruction manuals of a Samsung television of 2016, the one of a LG TV of 2017, the one of a Sony Bravia TV of 2018, the instruction manual of a Panasonic video device of 2012, the manual of a Sony DVD player of 2015, the instruction manual of a PlayStation 4 and the one of a Nintendo Switch together with their respective guides of use). As a translation memory is a specialized parallel corpus all those texts are inserted in the memory with their original version in English and its official translation into Spanish.

Taking into consideration the studies about representativity of corpora stated in the previous section of this paper, it is necessary to evaluate if the corpus employed in this paper is representative enough. Therefore, following the procedure used by Corpas Pastor and Seghiri (2010), the program Lexitrad's Recor has been used to evaluate if the corpus employed is valid enough to carry out the study.

The use of this application is quite intuitive. Only by entering the elements of the corpus in the entry data and a reference text in the output data allows the user to obtain interesting information about the representativeness of the input in the form of two graphics. These data are shown in the form of two graphs, a first graph that calculates the ratio types/tokens per document and a second one that calculates the same ratio per words.

Both graphs include two different calculations, one alphabetical one shown by the red line and one random one represented by the blue line.

As the corpus used for the study in this paper is much smaller than the Turicor used by Corpas Pastor and Seghiri (2010), the scale of the results is much smaller than this, but the results are not less important. Figures 7 to 10 show how the types/tokens ratio is drastically reduced from the fifth document and in all cases reaches a point of stability between the seventh and the tenth document, supporting the hypothesis raised by Biber (1995). However, a larger and more elaborate corpus would reduce the types/tokens ratio and would allow a better and more representative corpus.
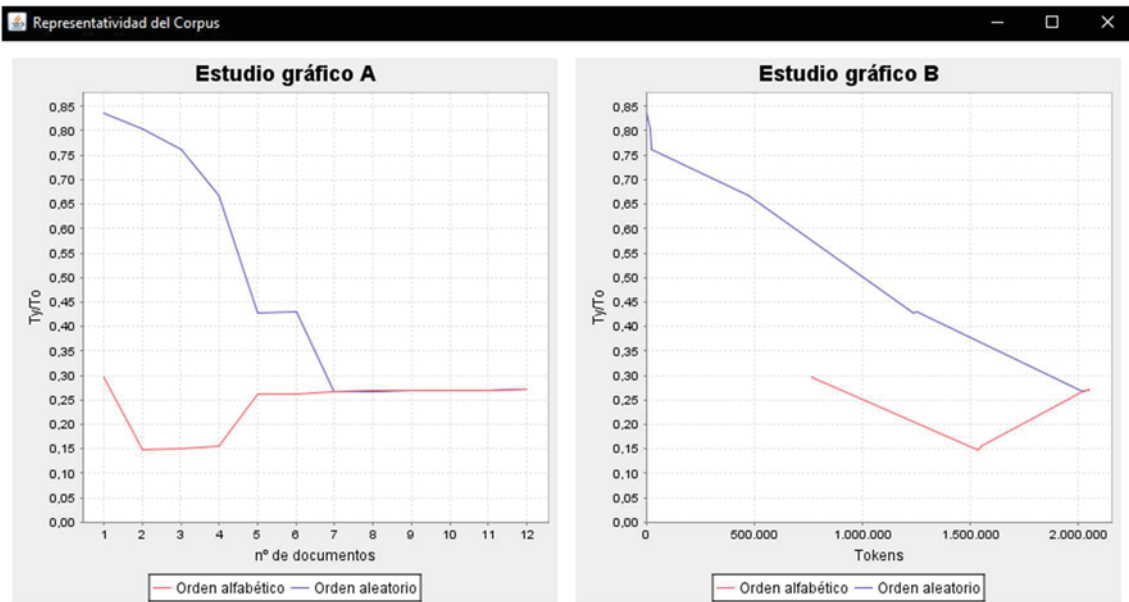
Figure 7: Representativeness of the English component (1-gram)
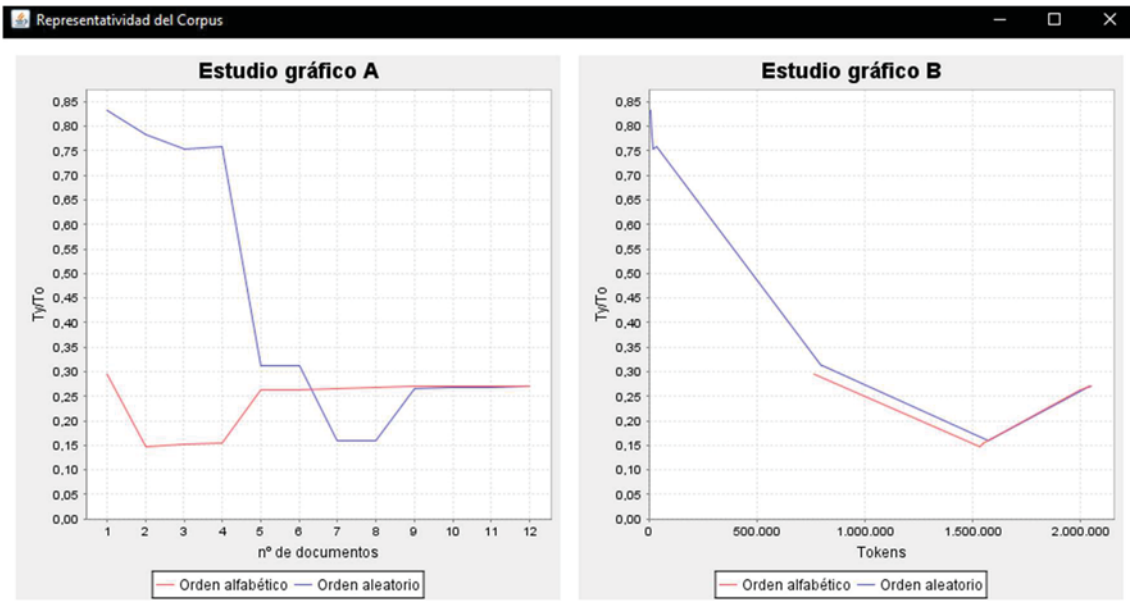
Figure 8: Representativeness of the Spanish component (1-gram)
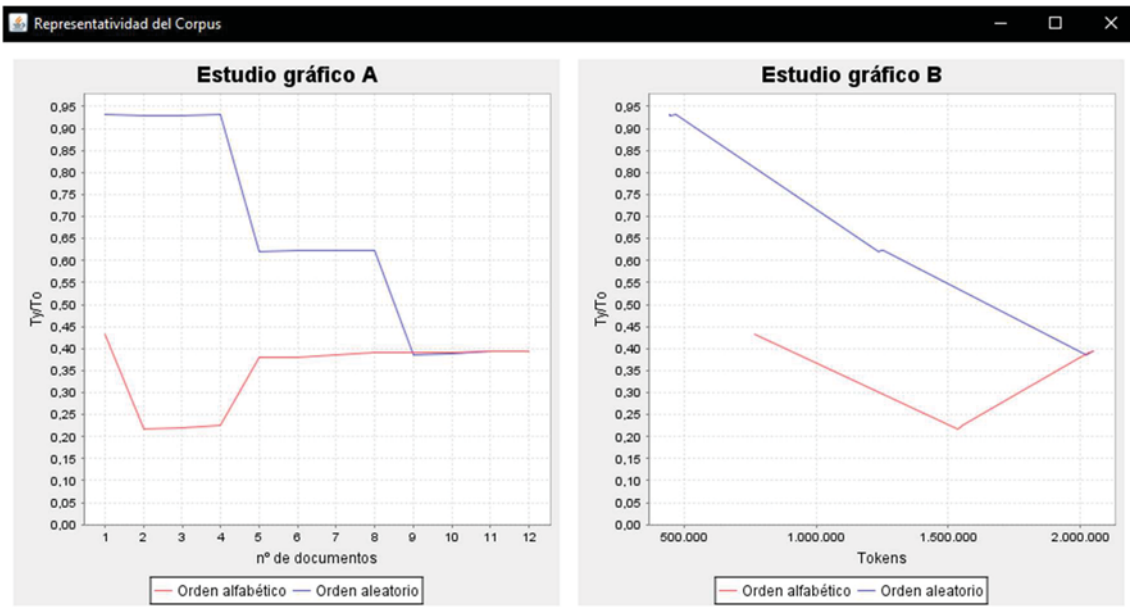

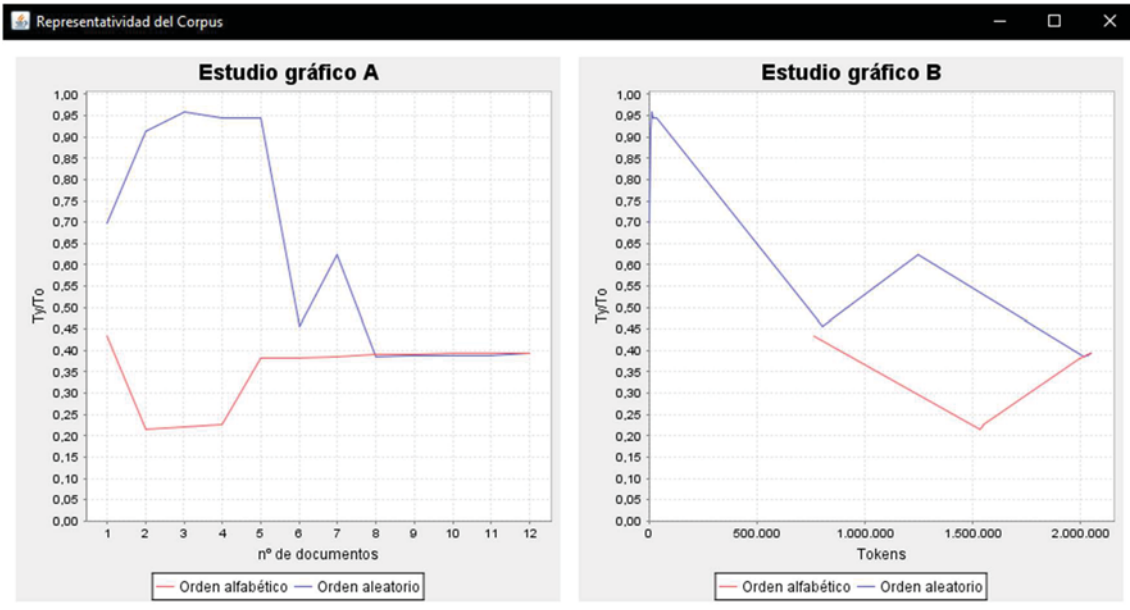*Figure 9: Representativeness of the English component (2-gram)*

Figure 10: Representativeness of the Spanish component (2-gram)

3.2. Methodology

Once the base translation memory is made and its representativeness checked, the final analysis of CAT tools can be done. This analysis is carried out by reviewing individually four different aspects of the various CAT tools. The methodology of the following sections consists of three main steps: individual analysis of each tool, comparative analysis and a final general conclusion.

One of the main aspects that a translator takes into account when deciding on a tool or another is the experience that this application has, the time it has been established in the market and the trust it conveys. It is therefore inevitable that the first aspect to be treated in the analysis of each tool is a brief historical introduction of each of the tools.

When the historical bases of the tool are established, the second aspect to study is the interface. At this point, we must take into account not only the external aspect of the application but the functionalities and unique features that the tool brings. For this, the analysis of this aspect will be carried out through the use of the translation memory elaborated previously in order to translate a new document of the same scope, an instruction manual of a Qled Samsung television of 2019, and see how that tool of work plays its role (see section 3.1).

17

Subsequently, once this in-depth analysis has been carried out, it is necessary to dedicate a third section, ease of use, in which it will be commented how the program interacts with the user and the facilities that this second person has to be able to carry out the task of translation.

Finally, the analysis of each tool has a final conclusion section in which a summary of the analyzed aspects will be made together with a small opinion of aspects to improve and other aspects that may have been leaved out.

Once the five tools of computer-assisted translation have been reviewed, this analysis has a final section in which, supported by a table in which the general aspects of each CAT tool are easily seen, these applications will be analyzed comparatively.

4. ANALYSIS AND EVALUATION

As stated in the previous sections this one will try to explore the main CAT tools (Trados, Wordfast, MemoQ, Déjà Vu and OmegaT) their functionalities and strong points as well as other aspects that could make unique the program in comparison with their rivals.

4.1. SDL Trados Studio

Developed by the German company Trados GmbH (TRAnslation and DOcumentation Software limited liability company) and distributed by the British company SDL plc (Software and Documentation Localization public limited company) SDL Trados Studio is the main Computer Assisted Translation tool in the world.

History

Trados started as a small company created by Jochen Hummel and Iko Knyphausen in 1984 in Germany who worked for IBM as an LSP (Language Service Provider). But it was not till 1990 when Trados launched their first own application, a terminology database called MultiTerm, marking the beginning of the current translation industry giant we know today.

It is important to remark that the tool known nowadays include many of the advancements made in the nineties, as the program T Align made by the computational linguist Matthias Heyn, known as WinAlign after 1997 with the acquisition of part of the company by Windows.

These advances are still important today and continue to be revised and updated every year, even more after the purchase of the company by the British SDL in 2005, in order to maintain the trust obtained by almost all professional translators.

Interface and functionalities

The first thing that comes to mind when opening the program for the first time is its incredible resemblance to the design of the current Microsoft Office suite with a clean and well divided interface in differentiated sections.

To carry out the analysis of the document to translate, in the following picture (Figure 11) it can be seen how the new project option contains a long series of menus that

reveal options and highlight elements of the program, such as the ability to translate into several languages at a time, or the capacity to import translation memories from other CAT tools and to work with terminology databases, whether they be external or handmade, in order to add a glossary of possible translations of every topic related word.
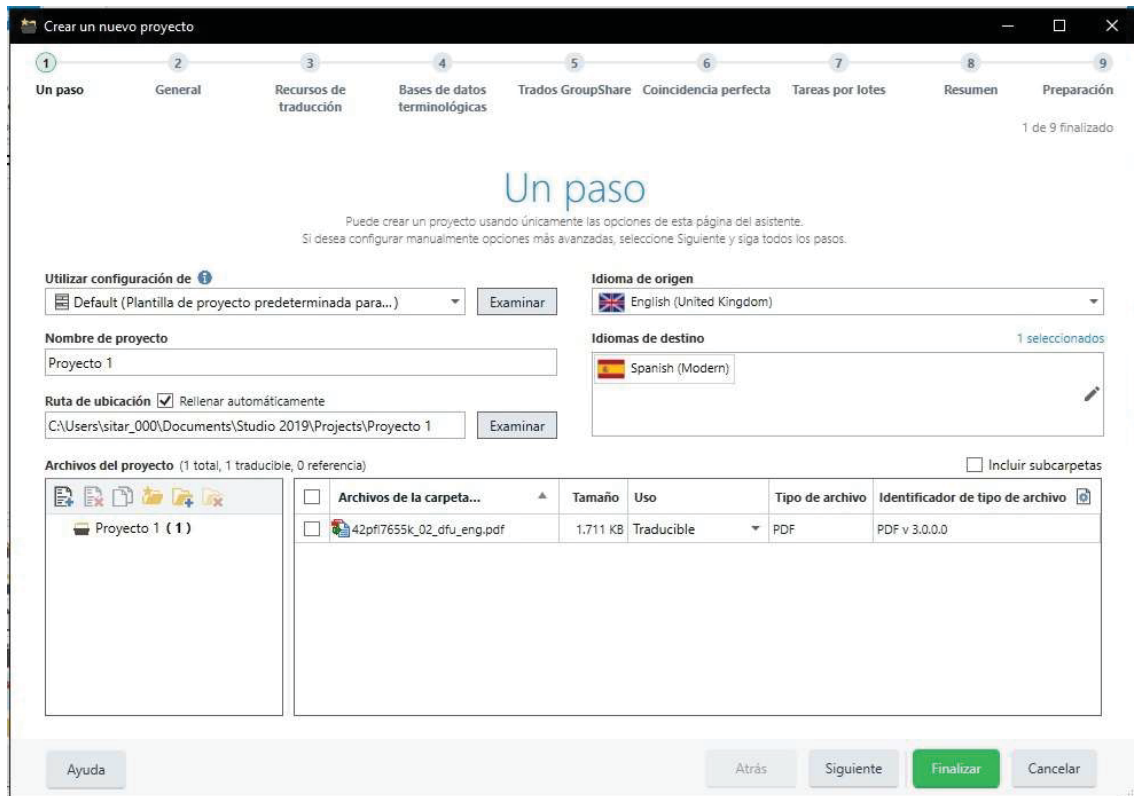

Figure 11: SDL Trados Studio new project screen

After the nine steps that the program requires when creating a new project, including the establishment of the above-mentioned translation memory, the progress bar shows clearly on figure 12 the first benefits of working with translation memories. Without having an exceptionally large or complex translation memory or using any terminology database, it can be seen how the computer-assisted translation tool has already translated automatically one fifth of the elements of the television manual to be translated.

Focusing more on the statistical data we can see how the translator's task has been reduced in 8,275 words of the 39,726 words included in the document, which means 20.83 percent of the words in the document. a slightly higher percentage can be seen in the number of translated characters, a 20.99 percent of the total, with 41,233 characters translated from the 196,377 characters included in the document. However, the most

striking data can be seen in the number of translated segments, of which Trados has translated using the small translation memory that we have offered 1,839 of the 4,947 segments contained in the instruction manual, that means 37.17 percent of the text segments, or in other words, more than a third of the segments contained in the document.

All these data taken from the experiential analysis will be evaluated in greater depth at the end of this section of the paper, after having shown the five programs to be analyzed.
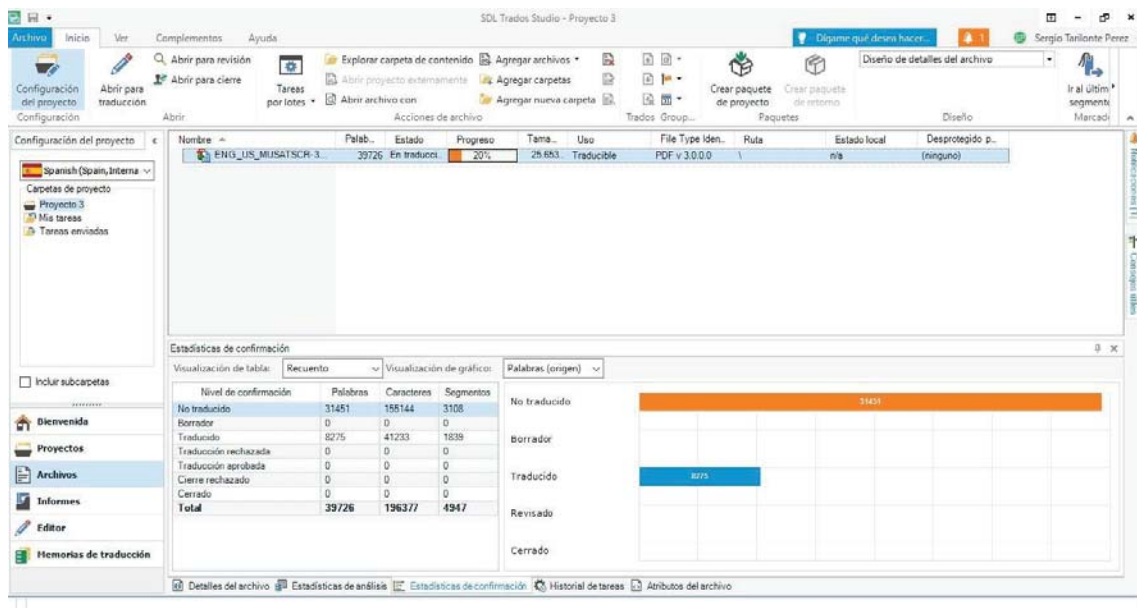
Figure 12: SDL Trados Studio project overview screen

In addition to this automatic translation of identical segments, the advantages of this assisted translation tool continue once the translator begins the translation of the remaining segments. After the translator is ready to translate a specific segment, if the program finds the translation memory a possible translation, or a segment that fits more than a previously marked percentage in the properties of the project (over seventy percent in the case of Figure 13), the system shows the user that possible segment or translation so that the only thing that needs to be done is to modify the translation until it fits perfectly with the translator's version.
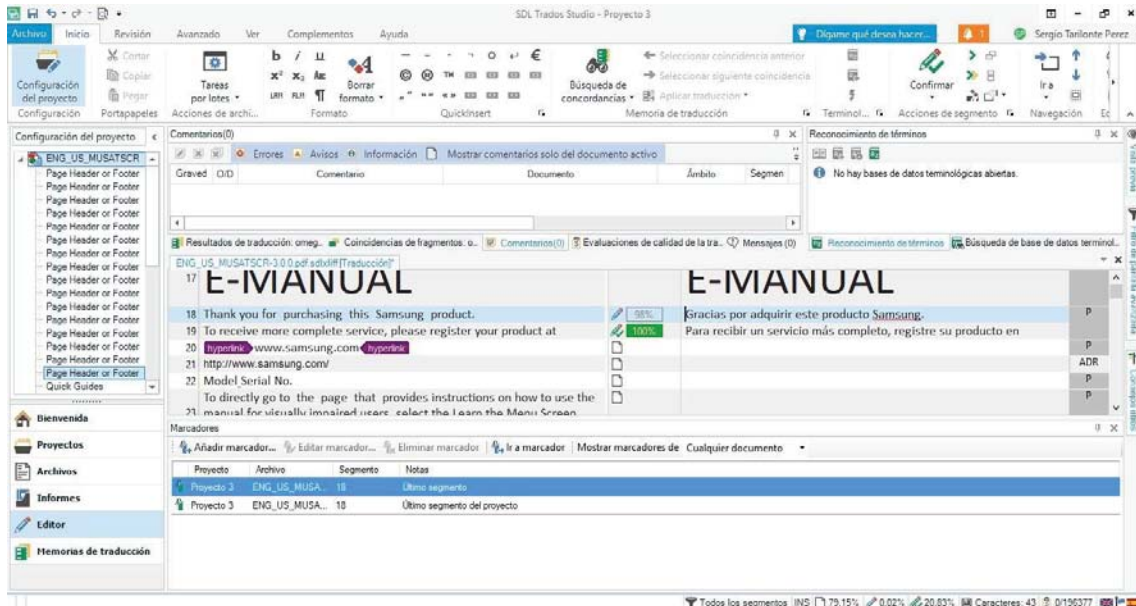
Figure 13: SDL Trados Studio segment translation screen

On the other hand, if the translator decides that the translation of a segment must be exactly the same as the original version, in the case, as in the case of the segment between hyperlink labels of figure 13, with a single right click the user can access a menu with multiple options, as shown in figure 14, such useful options as copying the source segment in the destination segment, looking for possible matches to that segment or adding comments if necessary.
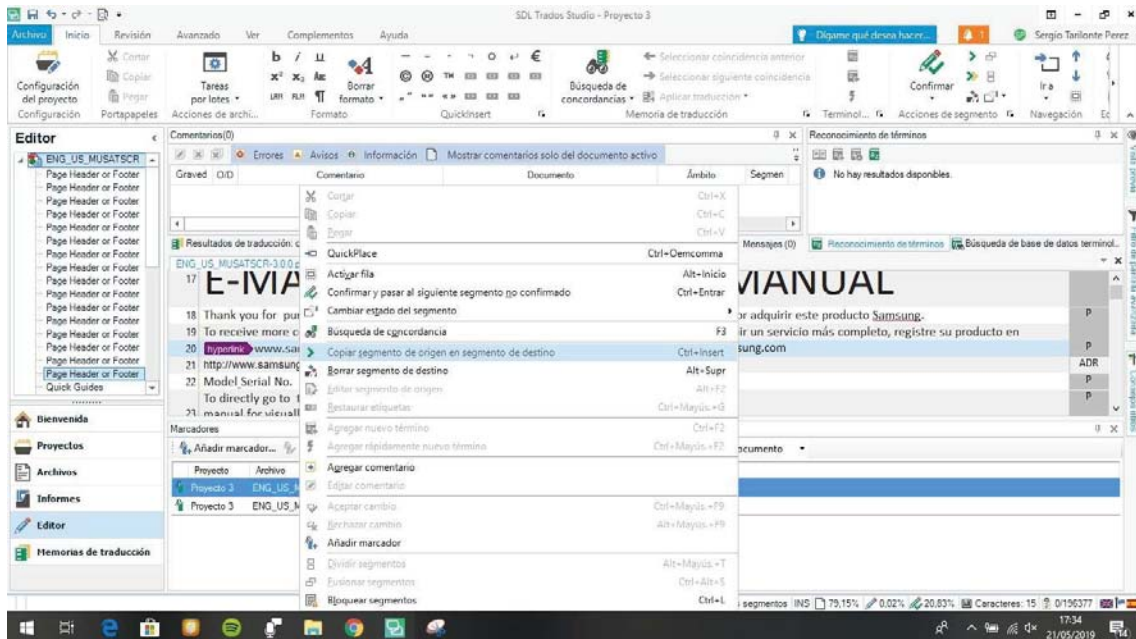
Figure 14: SDL Trados Studio segment translation drop-down menu

When talking about the interface and features offered by Trados Studio there are other small aspects that can make the difference when the translator chooses a CAT tool or another. Aspects such as good automatic segmentation of the program, which also allows modifying the segment by hand, the possibility of entering terminological databases as a dictionary always accessible to the translator and the ability to easily and comfortably edit the translation memory, such as can be seen in Figure 15, can become very relevant to the potential buyer of the product.
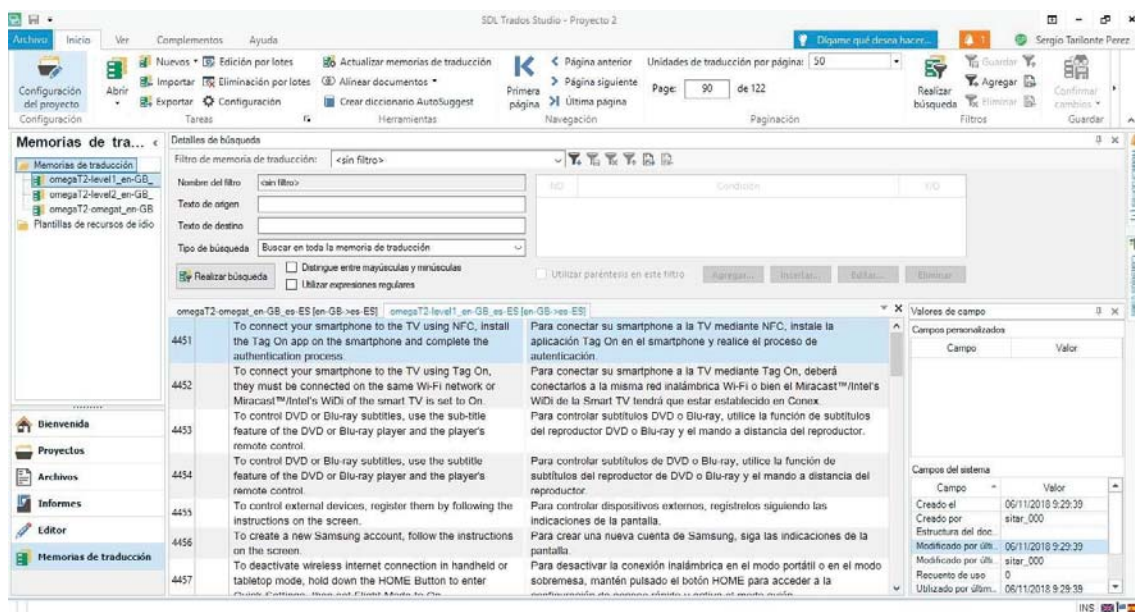

Figure 15: SDL Trados Studio translation memory screen

Ease of use

As mentioned above, the interface is very similar to the interface of the current Microsoft Office suite, which makes it easily accessible even for the least experienced user. In addition, the program includes a test document with a tutorial, shown in figure 16, for those users who do not know how to use the tool.

Learning throughout the use of the program is quite intuitive and with just a few sessions any translator who has just started using this program could easily navigate by avoiding certain advanced elements such as some of the programmable tasks included in the CAT tool.

In addition to this tutorial included in the program and the smooth learning curve of the program, Trados offers various training courses to which any user can sign up if he considers it necessary to strengthen his knowledge about the program.

However, although it is clear that Trados tries to combine all the tools that a translator may need, this can also be counterproductive because the excess of available options can overwhelm the user.
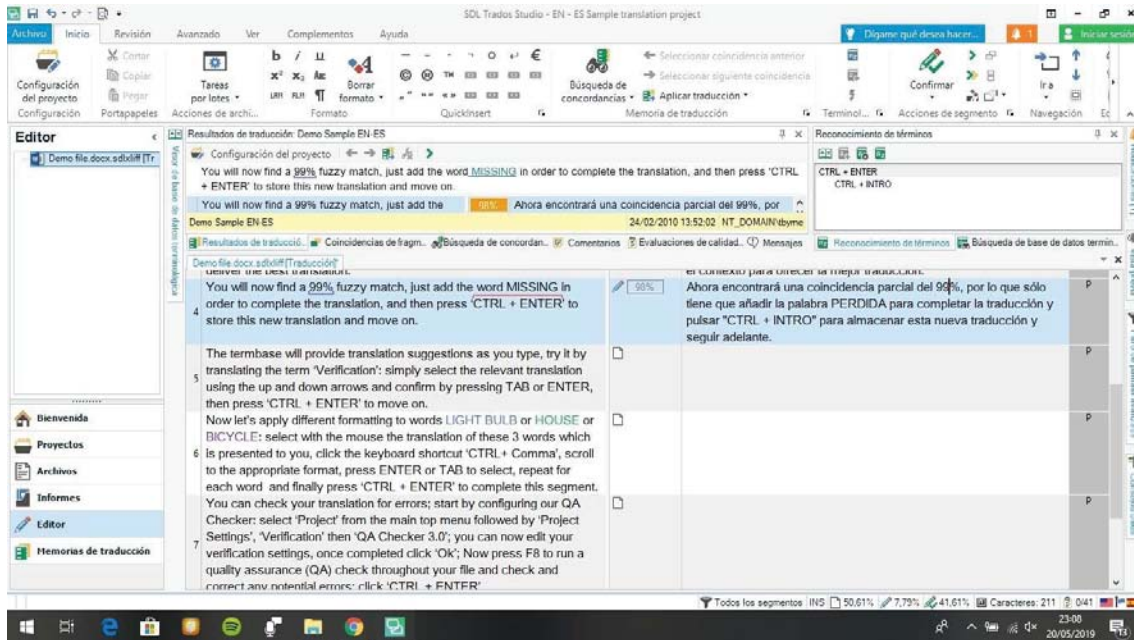

Figure 16: SDL Trados Studio demo file

Conclusions

To finish this small analysis of this computer-assisted translation tool, it is worth highlighting some things that could be improved in the face of future revisions of the program.

The first aspect to improve, and the one that has given most problems when preparing the analysis, is that when entering the source and destination languages includes an extensive list of languages with multiple varieties, such as international Spanish, modern Spanish, Spanish from the United States ... and if the user does not select the exact language of the translation memory as a translation language, it gives problems to the point of not reacting. This could easily be solved by allowing the installer language to those with which the translator knows that it will work and allow that if the user wants to add a new one afterwards; this can be added from some internal option of the program.

Another facet that draws the user's attention, this one much more understandable, is that the program only works with one translation memory at a time in order to avoid conflicts between them. A possible solution to this problem could be the possibility of

establishing a priority order within the memories. To give an example, if the translator works in a medical-scientific text and has a translation memory of medical texts and another one of scientific texts the program should go first to the medical one and then if it does not find any good match look in the memory of scientific texts.

On the other hand, it should be noted that Trados includes a store of installable applications as add-ons designed both by the company and by various users who have seen the needs of this tool, as shown in figure 17. However, the installation of these applications, including that of free add-ons, is restricted to users who have paid for the version of the program, which ranges between 99 and 2500 euros, which is a great disadvantage for that translator who wants to use the trial version in order to decide if it convinces him or not.

We can conclude that in spite of these slight drawbacks, SDL Trados Studio is a fantastic tool for any translator, regardless of their level, which makes it clear that today it has become a benchmark in computer-assisted translation.
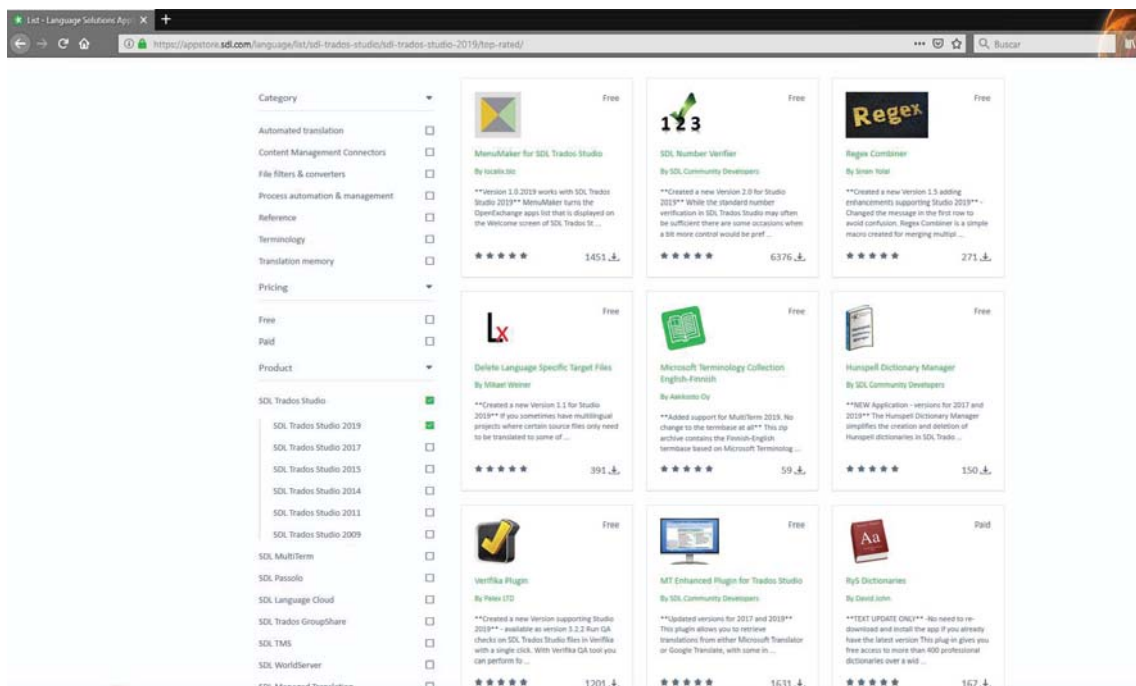

Figure 17: SDL Trados Studio appstore

4.2. Wordfast

Developed by the American company Wordfast LLC (limited liability company), Wordfast is today the main competence of Trados in the computer-assisted translation market.

History

The history of this program dates back to 1999, when the Frenchman Yves Champollion developed the application as a complement to the Microsoft Word program that turned out to be a cheaper alternative to the program already known at that time, Trados.

This tool quickly gained popularity because, until almost 2003 it was completely free. However, we cannot talk about Wordfast as an independent program until 2006 in which the owners of the translation service company TransPerfect, Philip Shawe and Elizabeth Elting, bought the rights to the add-on from its creator and started working on it. its independent version, which was not released until 2009.

Nowadays, Wordfast Studio has several applications that range from its own professional version to a quite limited free online version, through an add-on for Microsoft Word called Wordfast Classic that follows the model of its initial version. It also has its own server and a large and growing translation memory that is combined with other applications of the company.

Interface and functionalities

When opening the program, the first impression it conveys is how clear the interface is. If Trados seeks to become a multifunctional tool filled with many different options, Wordfast, in turn, tries to focus on the main tasks of a CAT tool.

Only with opening a new project to work with, it can be seen how Wordfast has taken a completely opposite route to its direct competition. As can be seen in figure 18 Wordfast has only one window in which the source language, the target languages, the files to be translated and the translation memories of the project are introduced. Apart from all those elements, this tool also allows to introduce or create a glossary and its opposite, a blacklist that allows the user to enter a list of translations to avoid.
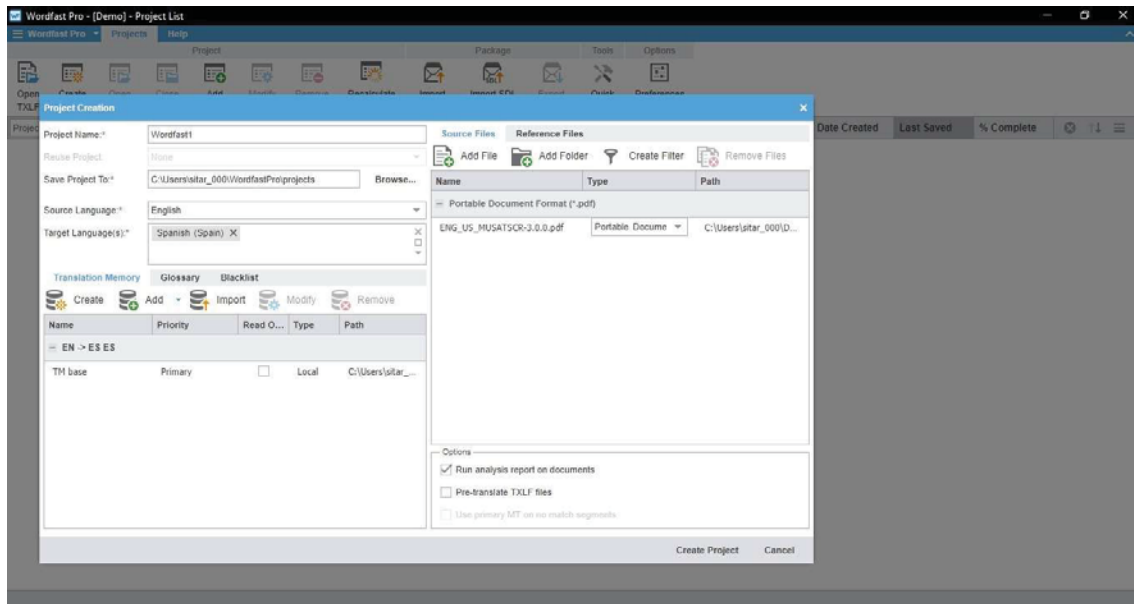
Figure 18: Wordfast Pro new project screen

Once all the necessary data have been established, it can be seen in figure 19 that Wordfast does not translate any segment at the beginning without being asked, even if it has a one hundred percent coincidence. However, this program includes the option *analyze*, where it shows the number of matches between the text to translate and the translation memory that the tool is able to find, and an automatic translation option, in which the tool automatically completes the segments that can be added and contain an agreement percentage equal to or higher than that indicated by the user. This option adds those segments that contain a greater percentage of matching among the options that it extracts from the translation memory, as can be seen in figure 20.
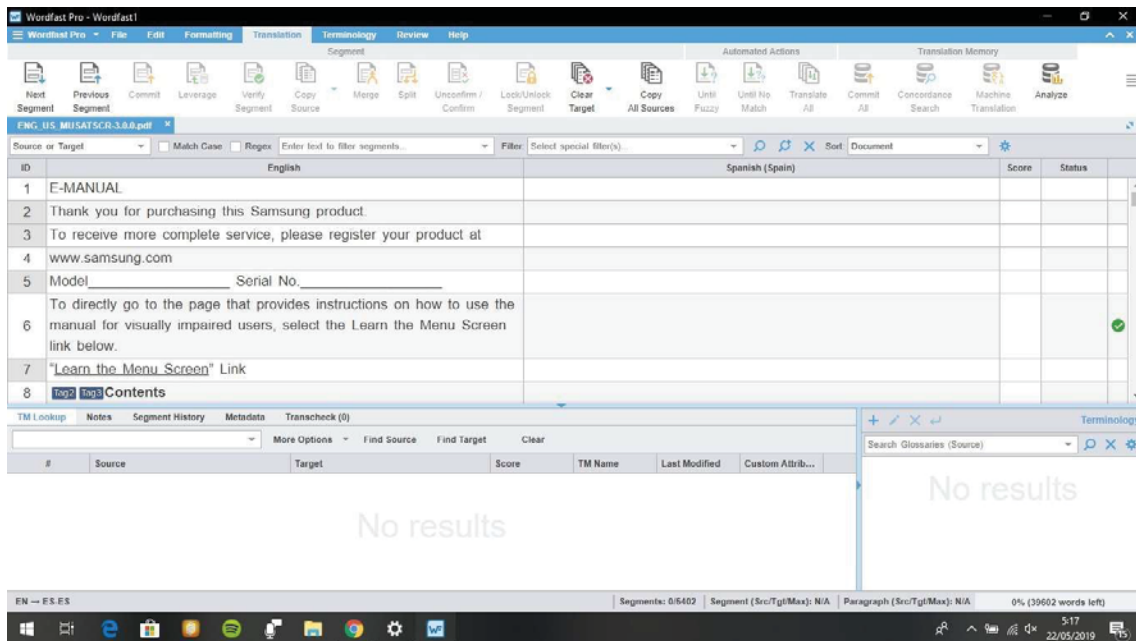
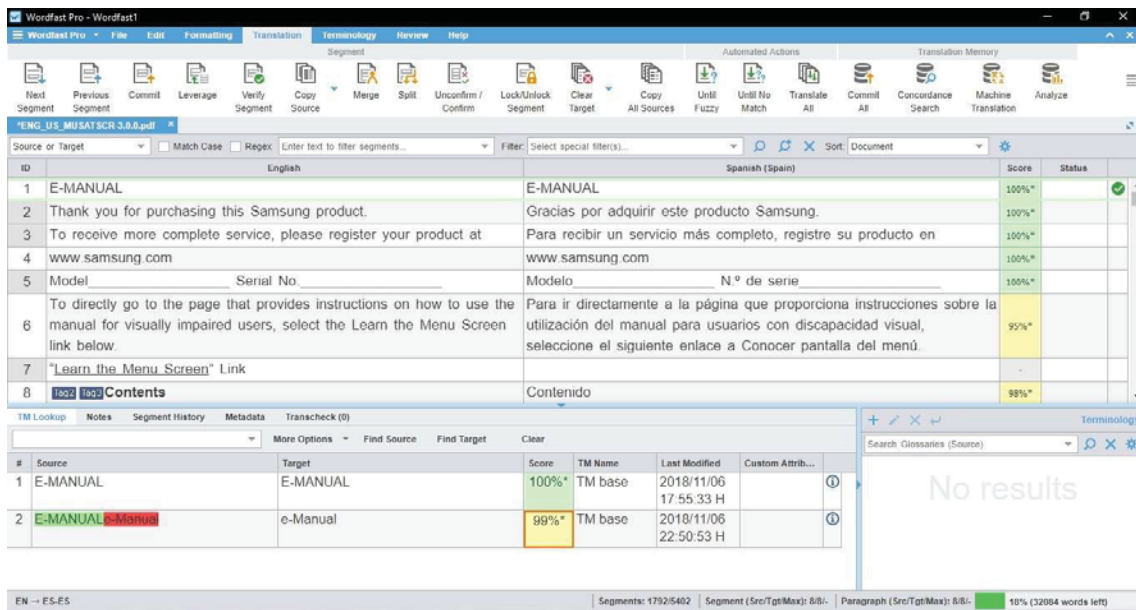*Figure 19: Wordfast Pro segment translation screen*

Figure 20: Wordfast Pro segment translation screen after automatic translation

In figure 20, in the lower part of the screen, it can also be seen that, when translating a segment, Wordfast provides a list of possible related translations. Every single translation is displayed with its percentage of possibility and, in the case of having introduced a terminological glossary, quick access to the words of the segment that are included in the glossary.

The statistical data of this application reveal interesting results. Analyzing it in the same way as the previous CAT tool, going only to the segments that produce a hundred percent of concordance results, it can be seen that of 39,602 words that the program finds, 7,503 produce a one hundred percent concordance,which supposes 18.94 percent of the words in the manual. On the other hand, the number of segments that produce one hundred percent of success is quite superior since of the 5,402 segments that the 1,741 program finds, they produce that percentage of success, 33.22 percent.

Wordfast does not provide specific results on the number of translated characters, however, this tool does include the number of concordances which value is under less than one hundred percent of matching. Analyzing the results higher than seventy percent can be seen as Wordfast has 3,584 translatable segments thanks to the translation memory, 66.34 percent of the total, and 25,141 words, 63.48 percent of the words throughout the document.

Report File ✕

Primary TM: none
Secondary TM: file://C:\Users\sitar_000\Documents\TM base

⊟ **Summary**

Match Types	Segments	Words	Percent	Tags
Leveraged	3986	28883	73	9299
Repetitions	898	6793	17	3390
TM Matches				
Context	0	0	0	0
100%	1741	7503	19	0
95%-99%	938	6775	17	3053
85%-94%	609	7596	19	3010
75%-84%	296	3267	8	1618
50%-74%	0	0	0	0
No Match	920	7668	20	4241
Ignored	0	0	0	0
Total	5402	39602	100	15312
Chars/Word	4.96			
Chars Total	196611			

Figure 21: Wordfast Pro analyze feature results screen

Ease of use

When talking about the ease of use of the program anyone can see that Wordfast is a slightly easier application than its main competitor, and therefore slightly easier to learn. Even in spite of not having, as in the case of its main competitor, a tutorial that explains the different functionalities of the tool, in just a few sessions any user can get to do the different elements that make up the program.

In addition, this computer-assisted translation tool includes its own wiki designed and modified only by its creators in which, thanks to the popularity of this kind of web pages, any user can easily locate the answers to any questions that may arise of the programs of this company.

Conclusions

As in the case of SDL Trados Studio, Wordfast is not exempt from aspects that could be improved in future revisions of the program.

One of the first aspects that draw the attention of any new user of the program is that the trial version of the program has an unlimited test time, but in return it reduces characteristics such as the number of segments or translation units of the translation memory to only 500, as well as the demo considerably reduces the maximum size of the document to be translated. This is a great inconvenience for the translator who wants to test in depth the characteristics of the program in order to opt for the before paying the 400 euros that is worth the full version nowadays.

On the other hand, another aspect that can frustrate the translator who decides to use this program is the long loading time that the tool has when performing any action that requires some time, such as opening the program, opening the project, analyze, translate automatically, search the document ... In our current society, we tend to want the work to be done as fast as possible and, although this kind of programs drastically reduce the time and effort of the translator's work, problems of this type can lead the user to opt for any other CAT tool that performs all these processes in a faster way.

In spite of this, Wordfast cannot be defined as a slow tool because, after the loading times mentioned before, this program works in a fast and dynamic way to stand up to any

of its competitors, being considered at present as the second-best computer-assisted translation tool.

4.3. MemoQ

Developed by the Hungarian company Kilgray Forditástechnológiai Kft. (Kilgray Translation Technologies limited liability company) MemoQ has been growing rapidly within the market of computer-assisted translation tools until today becoming one of the benchmarks in the market.

History

The origins of this computer-assisted translation tool date back to 2006 when Balázs Kis, István Lengyel and Gábor Ugray, whose surnames give rise to the name of the company, came together to lay the foundations for the first program version known today.

Today this CAT tool has three different formats depending on the user needs: a professional version for the basic user, a more advanced version designed for project managers and a version designed for translation servers.

Interface and functionalities

For any user who opens and sees the program interface for the first time, what is most striking is that, if Trados has an interface similar to that of the Microsoft Office suite, MemoQ takes that similarity a step further, becoming identical even in the file menus.

Once the user starts creating a new project, he can see that the project creation assistant is composed of two or three simple steps, which is considerably less than in the case of Trados. These steps ask for basic necessary things like the name of the project, the location, the input language and, unlike the previous CAT tools, the only target language, as can be seen in figure 22.
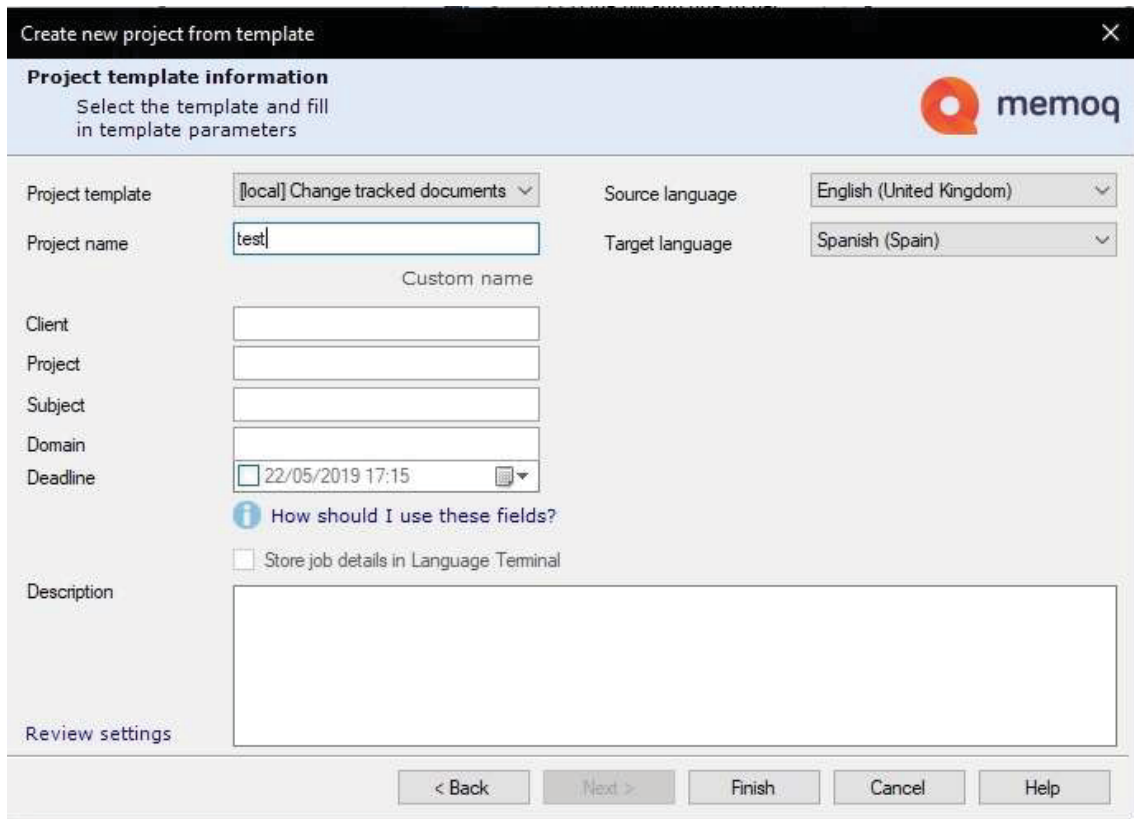
Figure 22: MemoQ new project screen

But, unlike the rest of CAT tools, the new project created is completely empty. Both the document to be translated and the terminological bases that the translator might want to use, such as translation memories must be created and imported manually, as shown in Figure 23. This means that the tool does not automatically translate the text without the user request, since the program may not be able to discern when they have finished putting memories and glossaries to the translation.
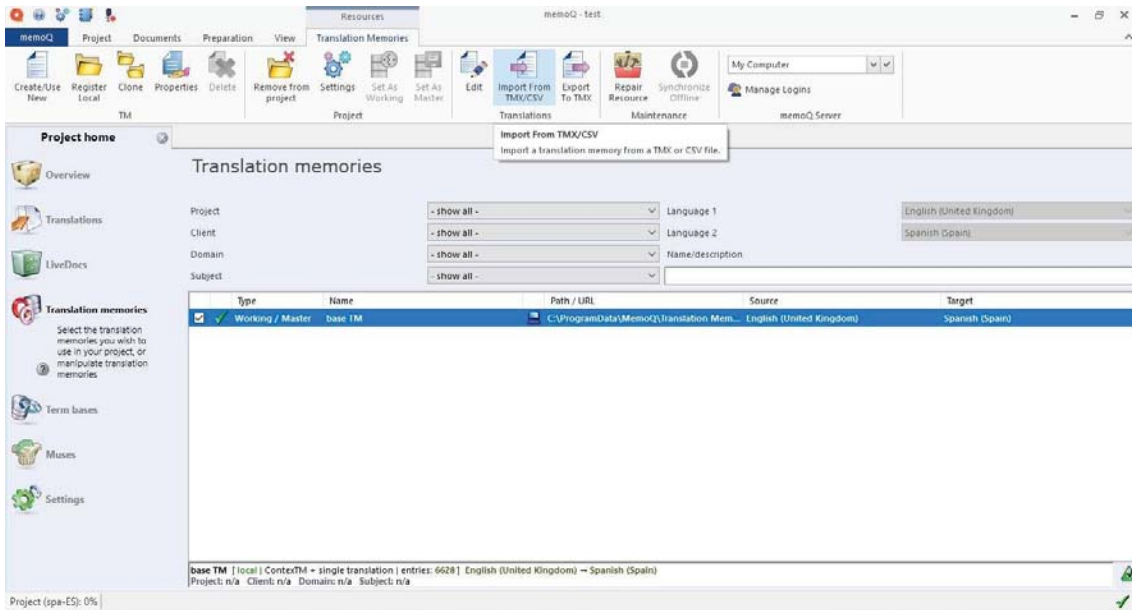
Figure 23: MemoQ translation memories screen

Once all the previous steps have been completed and the pretranslation has been carried out, the translator can start the translation, the verification or the editing, depending on how the segment is translated. As in the case of other CAT tools, MemoQ provides the user, as can be seen in the right part of figure 24, of the various options that the program has found in the translation memory along with data such as where the program is located the precision failures that do not give this one hundred percent matching.
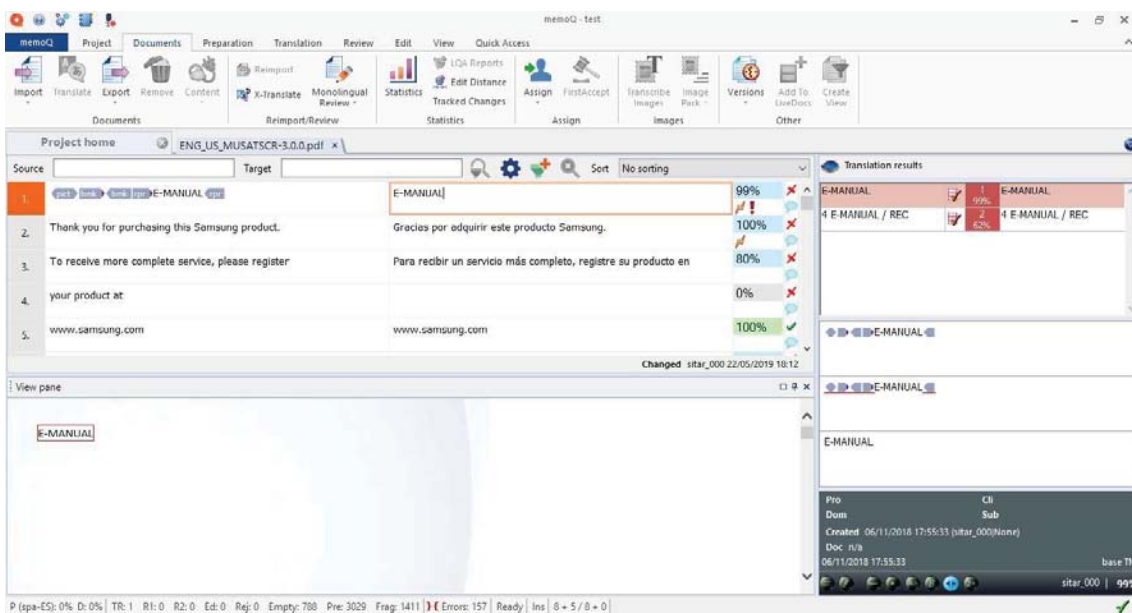
Figure 24: MemoQ segment translation screen

On the other hand, this tool includes many very eye-catching exclusive features.

A good example of these unique features of MemoQ is the management of errors, where the program marks the user the odd things it finds, such as extra spaces, missing labels, spelling errors… as shown in figure 25. These errors are marked near the matching percentage of the translation with a symbol of a lightning bolt, if these are minor errors, or that of an exclamation, if the program considers them serious errors.
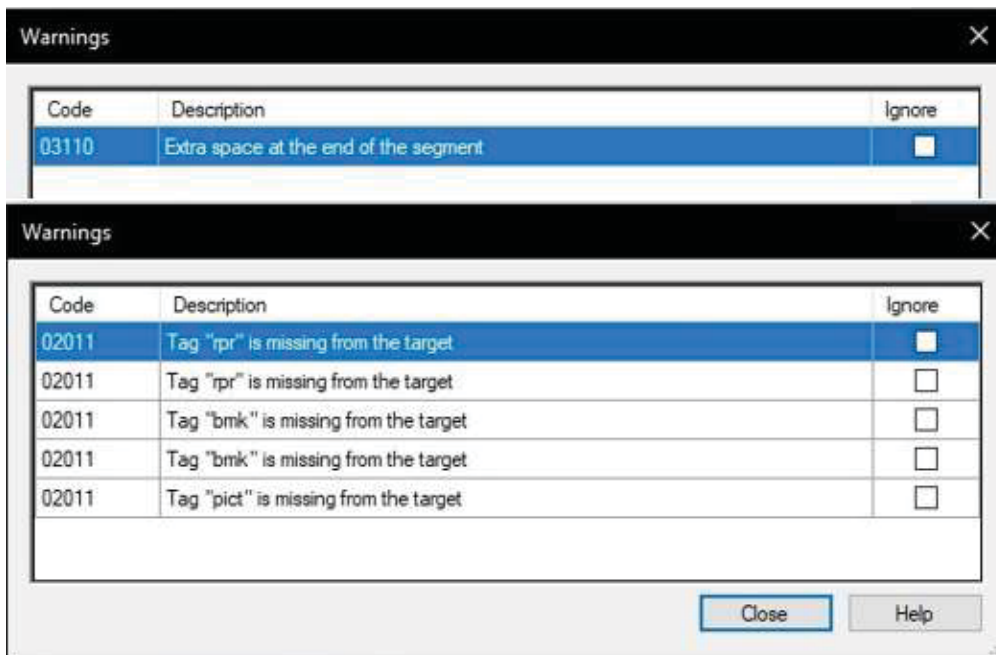

Figure 25: MemoQ error windows

Another useful feature of this program is the ability to separate and join segments during the automatic pretranslation process in order to find more and better matches in the procedure. This is very useful since, although the most common thing is that the program segments the texts from a dot to another or from one line to the next, sometimes these segments are cut in a different way than in the translation memory, so the program cannot find the most optimal translation in the translation memory.

As the last striking element of MemoQ, it should also be noted that this program allows the user to review the translations several times separately. This means that before a translation, the first reviewer can decide that it is correct, but a second reviewer can differ and mark it without altering the opinions of the first one. This is extremely effective when carrying out group translations, since not everyone has the same opinion about a translation.

The statistics of the translation of the text by using the translation memory are shown in figure 26. It can be seen that this CAT tool is a bit pickier because of the labels, but it does not produce negligible results. Of 40,315 words MemoQ translates with a hundred percent or more of matching 3,972 words, which is about 9.85 of all the words. Results similar to those of the character count, where 198.169 characters the program sees as perfectly translated 19.658, 9.91 percent of the total. Something higher is however the number of concordant segments where of 5,229 finds 794 segments, 15.18 percent of the total segments.

In addition to these results slightly lower than those of its competitors, MemoQ, just as Wordfast provides translation data with lower percentage of agreement. So, the user can see that if the program includes all the elements that exceed fifty percent agreement, which may only need to be postedited later, the results grow significantly. Doing this, MemoQ finds 28,955 translatable words, 71.82 percent of the total words of the manual, 142,339 characters, or as with the number of words a 71.82 percent of the total, and 3,569 segments, which is 68.25 percent of the total. All this without counting the data of repetitions that the statistic shows, since these can be or not translated.

Statistics for file(s) [eng-GB-spa-ES]ENG_US_MUSATSCR-3.0.0.pdf

Analysis

Scope   *ENG_US_MUSATSCR-3.0.0.pdf* (*"C:\Users\sitar_000\Desktop\ENG_US_MUSATSCR-3.0.0.pdf"*)
Resources   Every TM and corpus

Type	Segments	Source words	Source chars	Source tags	Percent
All	5229	40315	198169	25493	100,00
X-translated / double context	0	0	0	0	0,00
Repetition	1164	7831	38102	2701	19,42
101%	180	180	777	0	0,45
100%	614	3792	18881	5	9,41
95%-99%	519	3926	19834	4235	9,74
85%-94%	514	5504	26706	5593	13,65
75%-84%	587	6253	30403	3614	15,51
50%-74%	1155	9300	45738	6758	23,07
No match	496	3529	17728	2587	8,75

Figure 26: MemoQ statistics screen

Ease of use

When it comes to dealing with ease of use, it can be said that, as in the case of Trados, the design of the interface so similar to that of the globally used Microsoft Office suite makes the use of this tool easier for the user, making the user adapt quickly to the different functionalities of the program.

In addition, in order to make the learning curve of the program as smooth as possible, the program has on its website a large number of guides and courses related to the different products of the company, useful both for new users and for most experienced users in the world of CAT tools.

Conclusions

Although, according to the statistics shown, this program is slightly less efficient than the previous programs, MemoQ contains many elements that can turn it into the best option when it comes to choosing one of these tools, such as the ability to perform work and group corrections or the useful error detection system.

However, this program does not get rid of having things that can be improved, since things like the excessive slow pace of the program when it comes to reformatting a not-plain text into plain text, or the lack of options to translate a text into several languages at the same time can place it in a position worse than that of its competitors.

4.4. Déjà vu

Another of the most used CAT tools in the world, and one of eldest on the market is Déjà Vu developed by the French company Atril.

History

The origin of this tool goes back to 1993 when this kind of tools had just appeared. At that time its creator, the Spanish Emilio Benito, developed an application with Microsoft Word interface as a product to work out the needs of the company he was working for.

Even after the death of its creator in 2004, the CAT tool Déjà Vu continues to progress to this day, constantly including new versions and improvements that allow it to keep up with its competitors.

Interface and functionalities

As with the rest of the programs, the first aspect to analyze is what first catches the attention of any new user of the application. In the case of Déjà Vu, as it happens with

the Trados and MemoQ programs, it contains an interface that follows the designs of the Microsoft Office suite in order to be as user-friendly as possible.

When creating a new project, this program takes us, as in the case of Trados, through a series of windows, such as the one in figure 27, where a large amount of information is entered, for example: basic data of the project (name, folder, due date…), objectives, target languages, terminology databases or even automatic translation providers. It should be noted that although, in its free version, the system only allows the user to choose a target language, it can already be seen that in the professional version it is possible to choose more than one language at a time.
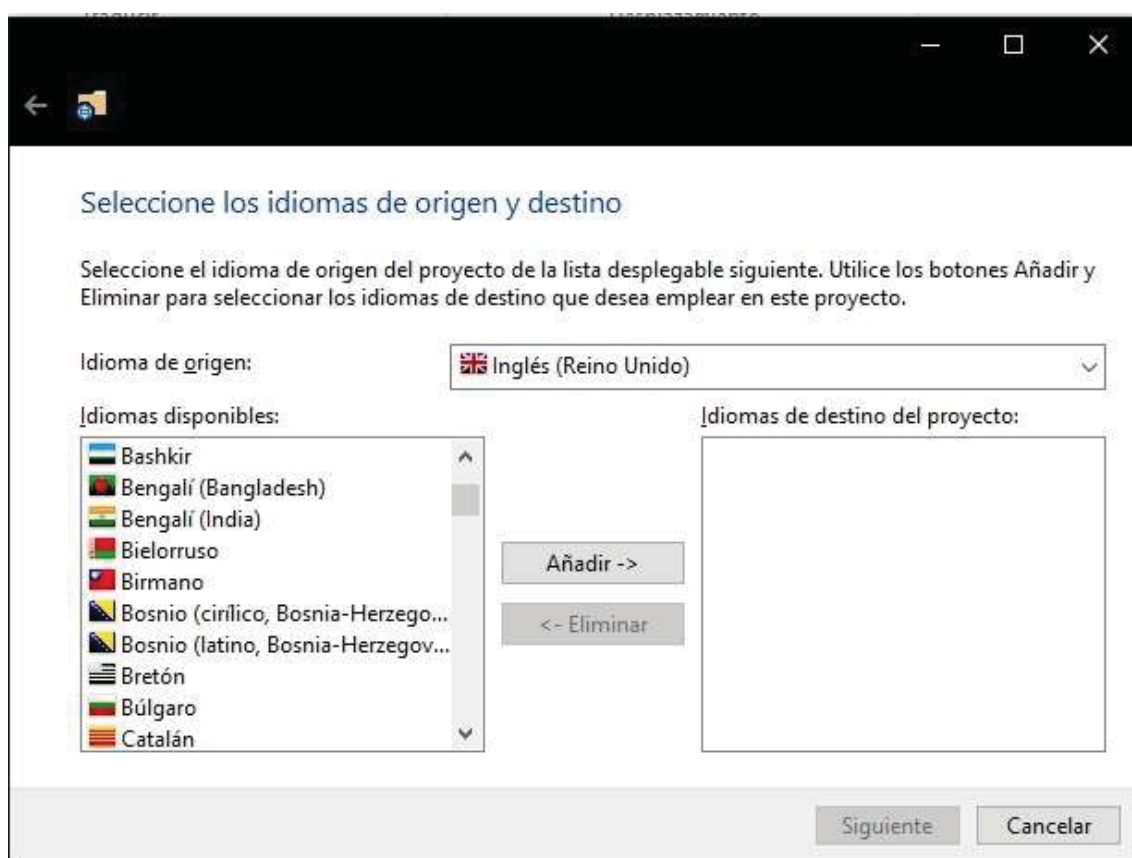
Figure 27: Déjà Vu new project screen

In contrast to the other CAT tools, even though the project creation gives the user the option of entering a translation memory, it must be in its own format and the process of introducing it is independent of the project itself. So, to be able to enter a translation memory first, it must be created as if it were a new project and, once the blank translation memory is created, import the data into this new file (only in .TMX format for the test version) using the option shown in Figure 28.
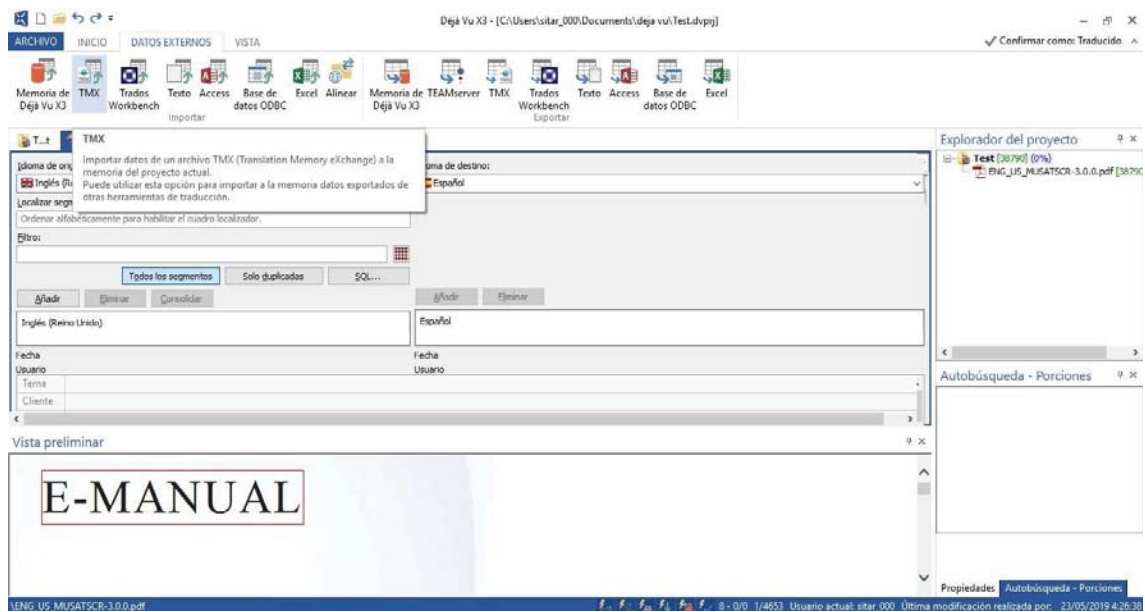
Figure 28: Déjà Vu external data menu

After the translation memory has been imported and added to the project, the translator's task can be started by using the pretranslation button located in the top menu of the program. This feature is full of options among which the user can access in order to obtain a more or less translated text with which to work; besides, it includes the possibility of automatically marking as valid those translations that give a hundred percent matching if the user is sure of its reliability.

As in the rest of tools of this type, the program shows according to the segment all the information about possible translations, as can be seen in the small box on the right in figure 29, as well as its source and percentage of probability, shown just below the box of possible translations.
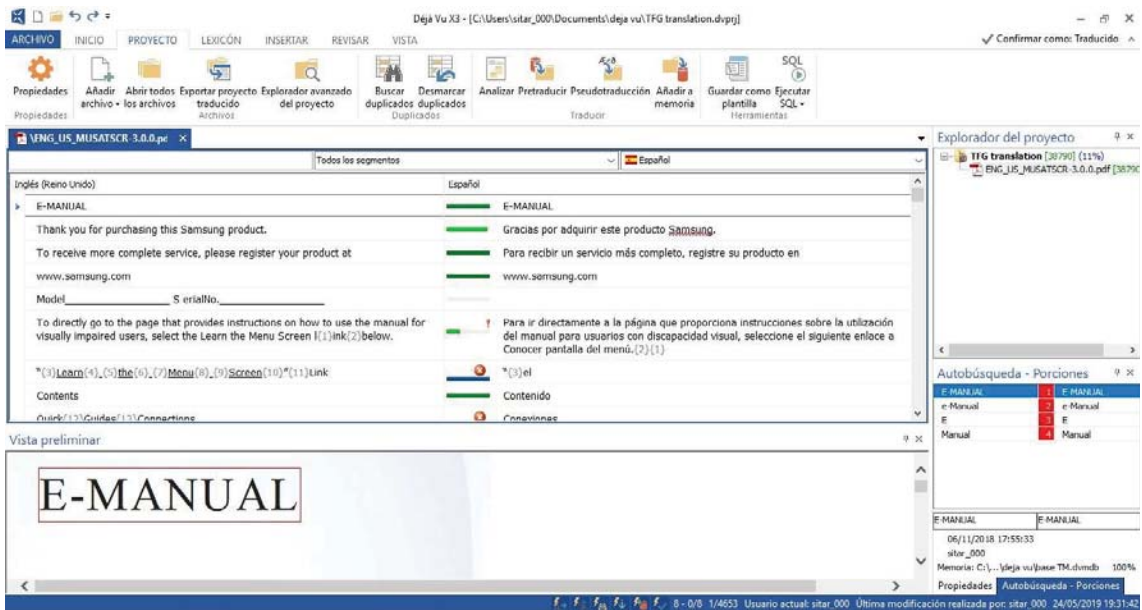
Figure 29: Déjà Vu segment translation screen

On the other hand, one of the latest additions to this program is a function to detect possible errors similar to the one contained in the MemoQ tool. Figure 30 clearly shows how the errors are marked with an exclamation mark or an x depending on how serious the program considers it and how the information referring to the error is shown by a pop-up text when the mouse passes through the segment with errors.
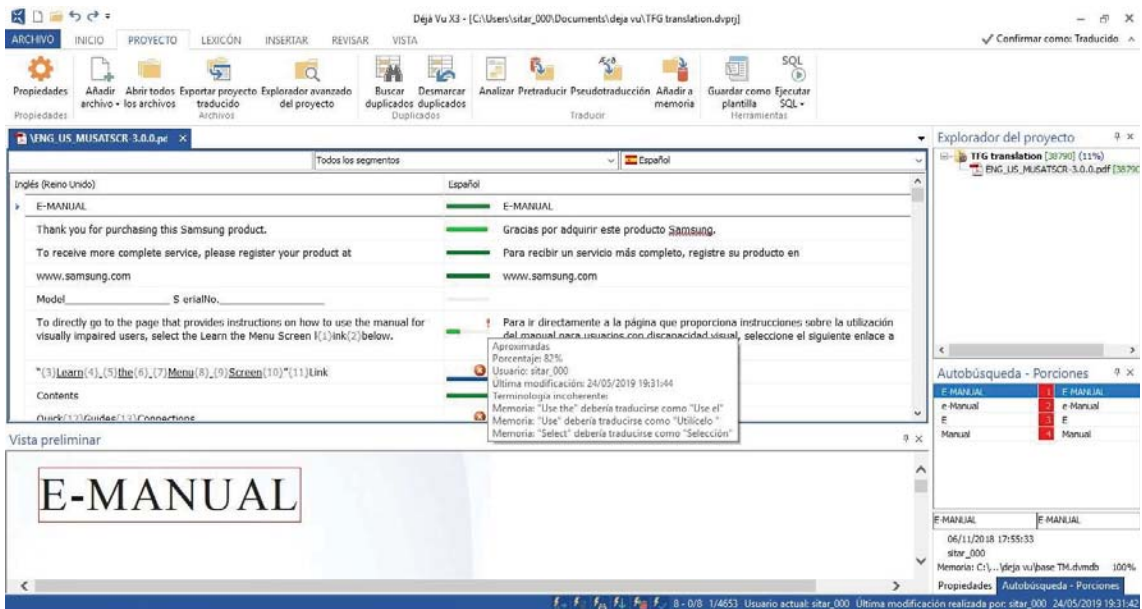

Figure 30: Déjà Vu segment translation screen error pop-up text

As in the rest of the computer-assisted translation tools, Déjà Vu has an analysis option in which all the numerical information regarding how the tool works is shown through the information provided to it (translation memories, terminological databases, automatic translators...). On this particular aspect, it should be noted that, of the five tools to analyze, this program has the best analysis tool due to its ability to customize the analysis based on the information that the user wants to obtain, however, this capacity also requires a longer processing time to be able to show all the desired data.

In order to obtain empirical results to compare the different CAT tools, figure 31 shows the results related to the analysis of pretranslation through the translation memory developed for this purpose, apart from other features of the analysis tool. Thus we can see that this tool translates with a hundred percent of matching 4,280 words of the 38,790 that it finds (11.03 percent), 25,937 characters of 235,582 (11 percent) and 1,055 segments of the 4,653 in which this program separates the document, a 22.67 percent of the total. These data alone indicate a better pretranslation capacity than the previous tool analyzed.

In addition, like the other tools analyzed above, with the exception of Trados, Déjà Vu also shows the results with a matching rate of less than one hundred percent. From these data it can be seen how, apart from repetitions that the program does not show if they have some type of percentage, the program is able to find similarity in 24,606 of the words (63.43 percent of the total), in 146,899 characters (62.35 percent) and 2,197 segments (47.21 percent). This allows the program to provide a pseudo-translation in almost 50 percent of the segments, limiting the translator's task in these segments to a simple revision or post-edition.
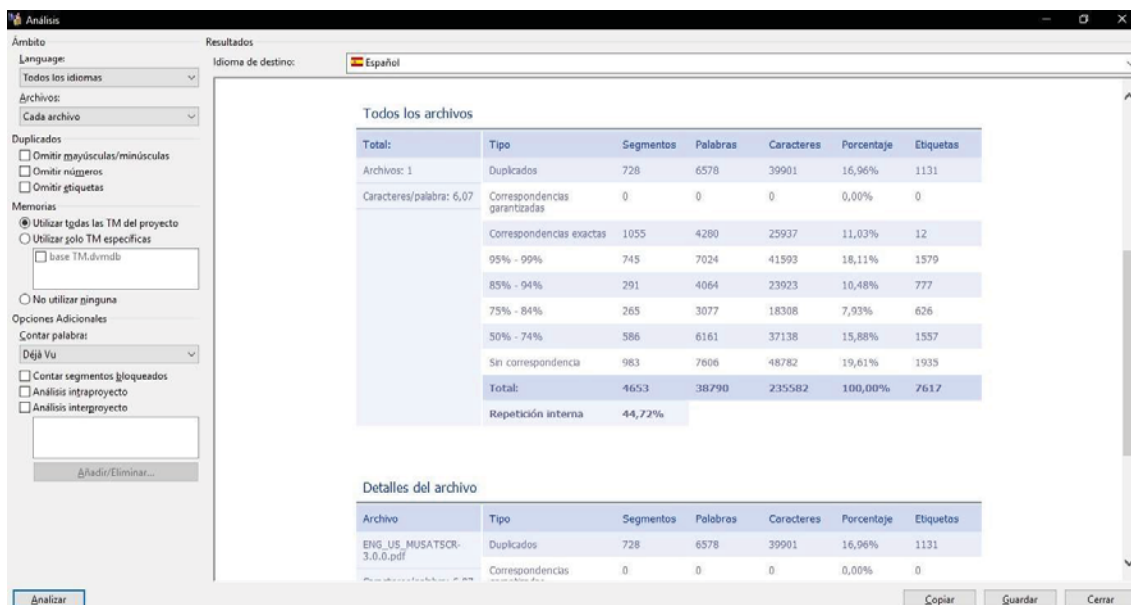
Figure 31: Déjà Vu analysis screen

Ease of use

In comparison with the rest of computer-assisted translation tools this program can also be considered as easy to use. This is because it also has two elements that make the learning process easy for the user.

On the one hand, as it has already been mentioned above, Déjà Vu has an interface that is tremendously similar to that of any application related to Microsoft Office to facilitate its use. On the other hand, Atril has on its page a series of manuals and tutorials that allow the solution of any doubt that may arise during the use of the program.

Conclusions

The analysis cannot be concluded without mentioning the small flaws that any user can find when using this program.

In the case of this program, these defects are mostly due to consumption issues, since, at least, when performing the analysis of the application, it could be observed the consumption of memory was far superior to the rest of the programs analyzed. This is a big problem for any user who works with old equipment because it may not be able to support so much memory and CPU load.

Moreover, the large number of add-ons in the program mean that the load times of functions such as the import of files into the software format or the analysis are much higher than usual.

However, none of these drawbacks are a reason not to consider this CAT tool as valid as any of the other tools.

4.5. OmegaT

The last of the tools to analyze is OmegaT. It is an open source application, so it does not belong to any specific company and any user can have their code to adapt it to their needs. As of today, this program is the preferred free computer-assisted translation tool among translators.

History

The history of this program goes back to the year 2000 when its designer, Keith Godfrey, developed the application as a tool for computer-assisted translation of plain texts. However, this first version was so limited that even its segmentation system limited itself to dividing the text into paragraphs.

Acros the years, this program has passed through the hands of various developers such as Maxym Mykhalchuk, Henry Pijffers, Didier Briel until it reaches current project leader Aaron Madlon-Kay.

Like most open source programs today, it is possible to find a verified and stable version on its website and a version in which developers add the different modifications suggested by users.

Interface and functionalities

It is difficult to analyze OmegaT in comparison with the aforementioned ones, because, as an open source program, it does not seek to compete with its rivals to obtain a piece of the market of computer assisted translation programs. This is noticeable from the first moment the user opens the application, since the first one is shown as a simple-looking program, more similar to Notepad than to Microsoft Word.

When starting a new project (figure 32) the user can also appreciate the simplicity of the program, choosing only the directory where the project will start, the source

language and the only target language and the folders where the different files will be found, both the folders where the files needed for the translation task are located, and the folders that will contain the files created. However, the image below also shows that, in most cases, what it is introduced is not the files themselves, but the folders that contain those files, so those ones must be added within the selected folders externally to the program.
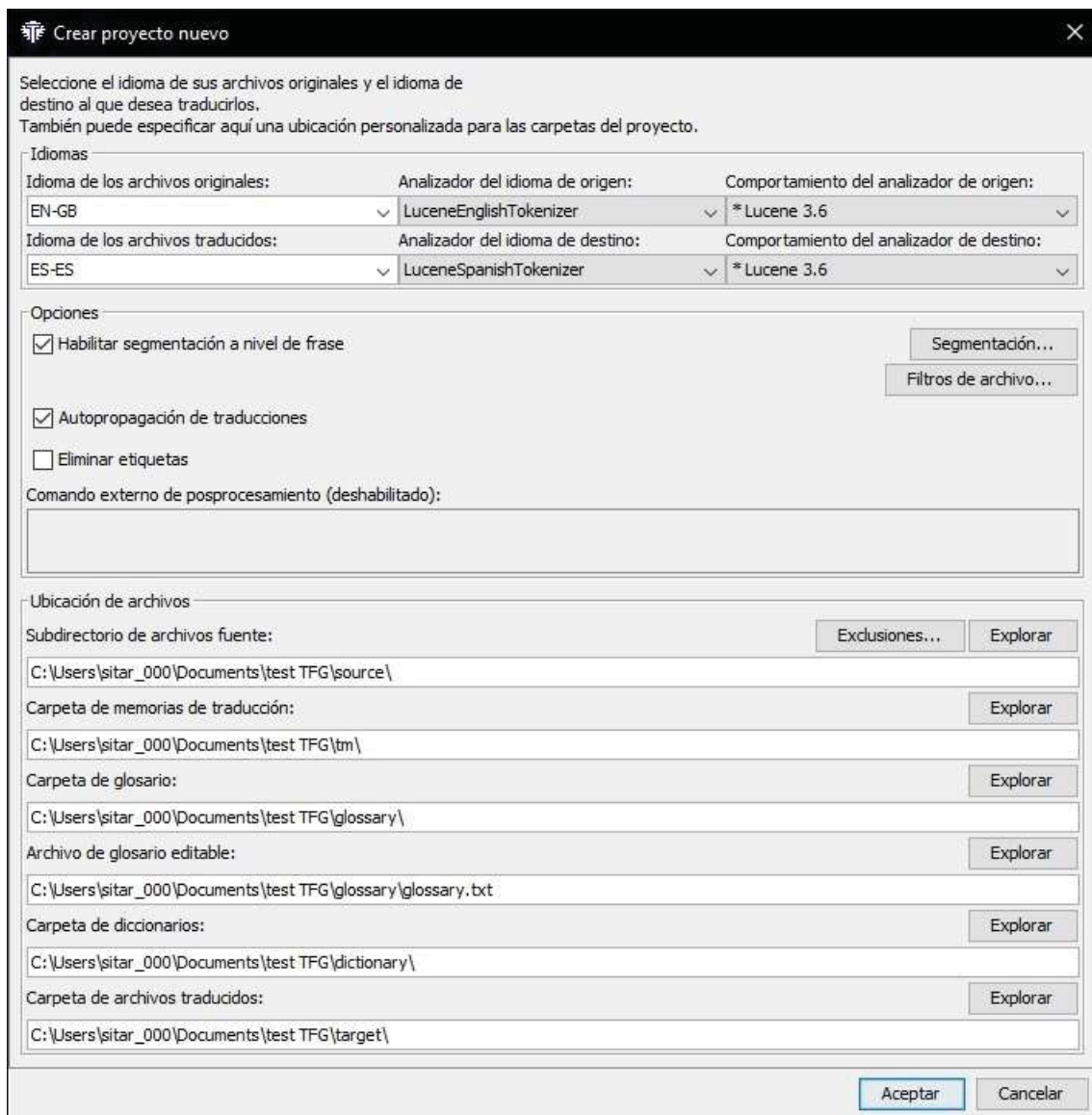

Figure 32: OmegaT new project screen

Once the file, or files, that the translator wants to translate have been imported and selected to start translating, the user can check again the weaknesses of this program compared to its payment counterparts. Figure 33 shows how the program, in its stable version without modifications, has many more problems to segment the texts and does

not have a button that allows to pre-translate the text automatically. In turn, it can also be observed that apart from the translation suggestions, the program also suggests self-generated phrases through the translation memory and the various add-ons that are added to the program, such as those that can be seen in figure 34 that are introduced in the base version of the tool.
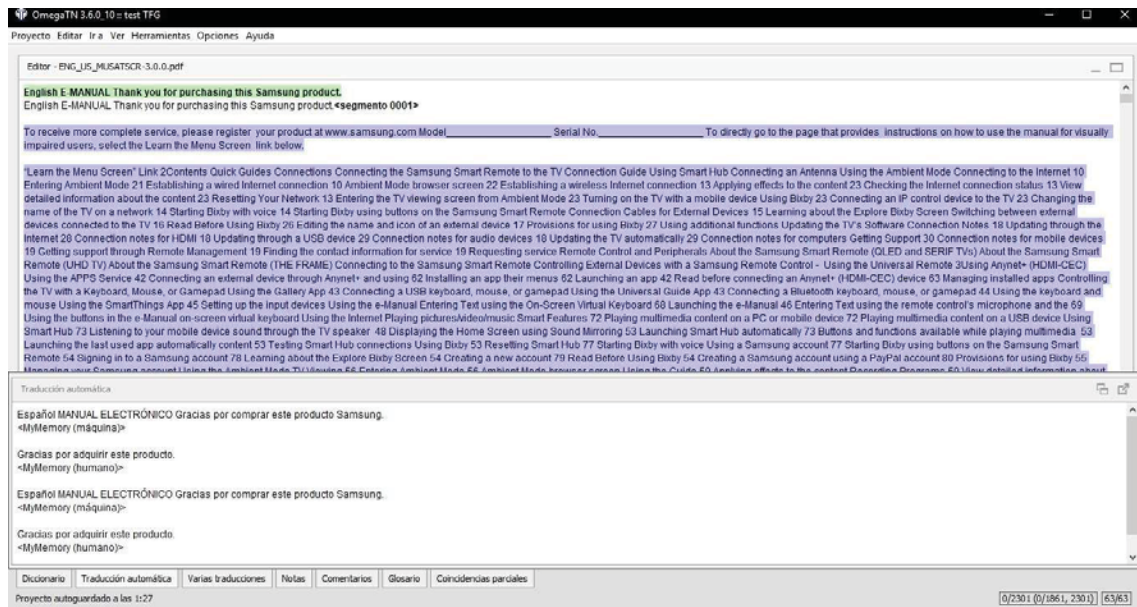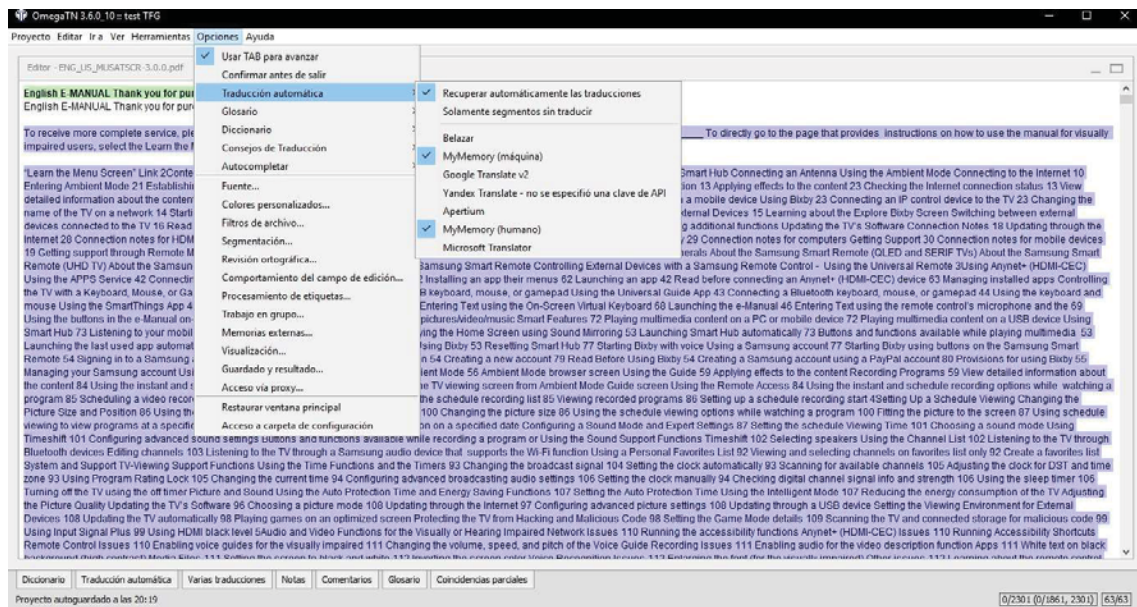

Figure 33: OmegaT segment translation screen

Figure 34: OmegaT automatic translation drop-down menu

As for the statistics system of the program the user can see how, despite the weaknesses of this open source application, they have nothing to envy to the statistics of many of the payment programs, as seen in figure 35. In this way the user can see without

venturing into the entire translation process as of the 40,018 words detected by the program 4,783 can be translated with almost one hundred percent matching (11.95 percent), of the 198,309 characters the program detects a translation between ninety and hundred and one hundred percent in 22,895 (11.54 percent) and, in the case of the segments, 363 of the 2,301 in which the document divides, 15.77 percent of the total file.

OmegaT, in addition to the maximum match statistics, also exposes similar results to those of applications such as MemoQ, giving other minor coincidence percentage that appear in the suggestions when performing the translation task. So, we can see how, in total, without counting possible repetitions, the percentages change to 60.83 percent in the case of words and 60.69 percent of the characters (24,344 words and 120,356 characters respectively) and as the total of segments in Those that the program can anticipate information go up to 1,497 segments, a 65.05 percent.

	Segmentos	Palabras	Caracteres (sin espacios)	Caracteres (con espacios)
Repeticiones:	440	5.484	27.026	32.518
Coincidencia exacta:	0	0	0	0
95%-100%:	363	4.783	22.895	27.485
85%-94%:	381	5.467	26.896	32.362
75%-84%:	211	3.053	14.778	17.857
50%-74%:	542	11.041	55.787	67.299
Sin coincidencia:	364	10.190	50.927	62.273
Total:	2.301	40.018	198.309	239.794

Figure 35: OmegaT statistics window

Ease of use

Speaking in terms of ease of use, OmegaT, despite being probably the simplest of the five CAT tools, can also prove to be the most complex to use for a new user due to its simpler interface and less oriented to facilitate the work of the user. However, like the rest of its rivals, it also has a series of tutorials, both in video and manuals to facilitate the task of anyone who may have doubts about how any feature of the program works.

In addition to the latter, as it is an open source application, any translator with sufficient programming knowledge can take the source code of the program and reapply it according to their needs, giving this program a range of possibilities that can be even greater than the of its competitors.

Conclusions

As a small conclusion to the analysis of this tool, it is important to re-emphasize that this program may be the least developed of those analyzed, by not competing for more benefits than its competitors, but its open source makes it a blank canvas for a person with sufficient programming knowledge. Furthermore, whether the user is experienced in programming or not, has a series of add-ons on the web page, as shown in figure 36, developed in order to add new or better features.

Figure 36: OmegaT external resources

4.6. Evaluation of CAT Tools

When making a comparative analysis of all the previous computer-assisted translation tools, this analysis can be done in two ways. The first of the ways to analyze the different applications is through an analysis of the different features of the program. The second tries to make a comparison based on the data obtained from the empirical analysis of how each tool interacts with the translation memory previously prepared.

Feature evaluation

The first thing that should be said when comparing the different applications analyzed is that, as can be seen in table 1 (located on page 46), none of the programs is completely superior to the others.

On the one hand, we can see how most of these tools are quite easy to use and master, except for OmegaT, which in turn may cost a bit more to be fluent in it. Analyzing the table, it can be deduced that, apart from the interface, as mentioned in the individual analysis, one of the reasons for this added complexity may be the absence of a pretranslation option, unlike in the case of Trados, where the program even translates the different segments automatically each time the tool has enough information without the need for the user to press any button.

Apart from this, the user can see how the absence of being able to translate more than one language at a time, like MemoQ and the free version of Déjà Vu, leaves these applications a step below the other two programs, SDL Trados and Wordfast.

However, to compensate for this lack, it can also be observed how both MemoQ and Déjà Vu autogenerate a glossary that supports the translation suggestion system, with its different percentages of matching, as well as a detector of possible translation errors.

In addition to all these aspects, and others not mentioned in the present paper, other minor features can also be highlighted, such as the case of aids to new users where it can be seen that most of the programs offer tutorials and guides but that Wordfast is the only one that uses a wiki as a method of tutelage. On the other hand, it is also worth noting that OmegaT has the advantage of having video tutorials as a reinforcement in its manuals section.

Another minor aspect, which may be relevant for some, is that Déjà Vu and Wordfast are left behind in terms of speed; the first one because of its slowness when performing tasks, and the second one because of its high memory and CPU consumption, as it is commented in the individual analysis.

Finally, although a larger study could be spoken of other very influential aspects such as how the terminology bases interact or even the price of the program, it is worth noting the capacity to improve the program of both Trados and OmegaT through the inclusion of add -on downloadable in their respective web pages that give these applications an additional value with respect to their competition.

	Ease of use	Pretranslation	Target languages at a time	Translation suggestion	Error Detection	Aids for new users	Speed	Add-ons
SDL Trados Studio	Easy	Automatically add segments with more than the wanted percentage	One or more	Automatically adds best option. Shows other options below	No	In app tutorial, training courses and manuals	Fast	Yes, on their online shop
Wordfast	Easy	Yes, with a button	One or more	Sentence suggestions	No	Wiki	Slow	Not officially
MemoQ	Easy	Yes, with a button	One	Sentence suggestions plus autogenerated glossary	Yes	Guides and courses	Fast	Not officially
Déjà Vu	Easy	Yes, with a button	One, more on pay versions	Sentence suggestions plus autogenerated glossary	Yes	Manuals and tutorials	Intermediate due to its heavy memory use	Not officially
OmegaT	Mediocre	No, only suggestions for the user	One	Sentence suggestions	No	Manuals and video tutorials	Fast	Yes

Table 1: Feature evaluation

Empirical evaluation

Another crucial aspect when comparing all these tools is how they perform their task. To prove that, there is no better way than submitting the same task to all the software and obtain numerical results to establish the evaluation of this data. In this paper, the text that has been given is, in particular, a manual in PDF format (non-plain text) to be able to get additional data on how the tools behave at the time of importing and segmenting something slightly more complex for the program.

Surprisingly, the results of the five programs could not be more different in all aspects, as shown in table 2 (on page 49), leading to some conclusions such as those cited below.

Just looking at the data of how the words and the total segments of the documents are imported and detected, it is observed that Déjà Vu obtains almost forty thousand characters more than the others, while at the same time, it is the one that finds fewer words in the document; MemoQ, on the other hand is, in this second aspect, the one with the most words with a result slightly superior to the rest.

When it comes to taking all the information and dividing it into segments to translate, OmegaT produces more than two thousand segments less than the rest, being in this case Wordfast the one that segments the text into more parts.

After the input data is analyzed, the user observes the data on how these tools perform the task of pretranslation; in this sense, Trados, honoring its position in the global market, is the tool that gets best initial results with a hundred percent matching, despite the fact that, unfortunately, in its statistics, it does not show any data on coincidences less than one hundred percent.

On the other hand, it can be observed how Wordfast offers results with a hundred percent matching equally worthy in comparison with the rest of its competitors; it can be seen how the next one of the list does not reach 23 percent of maximum matching when it comes to translating segments, ten percent less than Wordfast, and how these competitors stay between ten and twelve percent when it comes to perfectly translating words and characters, around eight percent less than Wordfast.

MemoQ on the other hand gives a higher percentage of total translation suggestions (100% and less) but it is the one with the fewest perfect hits, possibly due to how it manages the labels

All this information changes if we add the results with lower matching index, of which the computer-assisted translation tool will give a translation suggestion that is not so perfect but may only need a slight edition.

In this case, MemoQ stands out over the rest, giving results of 68 percent of the segments and almost 72 percent of the words and characters, or more if we take into account that none of the programs say whether repeated data elements produce coincidence or no. The most unfavorable result in this aspect would be for Déjà Vu that does not manage to give translation suggestions nor to 50 percent of the segments in which it divides the document.

| | Total | | | Maximum matching | | | | | | Match greater than 50 percent* | | | | | |
| | Number | | | Number | | | Percentage | | | Number | | | Percentage | | |
	Characters	Words	Segments	Characters	Words	Segments	Characters	Words	Segments	Characters	Words	Segments	Characters	Words	Segments
SDL Trados Studio	196,377	39,726	4,947	41,233	8,275	1,839	20.99	20.83	37.17	-	-	-	-	-	-
Wordfast	196,611	39,602	5,402	-	7,503	1,741	-	18.94	33.22	-	25,141	3,584	-	63.48	66.34
MemoQ	198,169	40,315	5,229	19,658	3,972	794	9.91	9.85	15.18	142,399	28,955	3,569	71.82	71.82	68.25
Déjà Vu	235,582	38,790	4,653	25,937	4,280	1,055	11.00	11.03	22.67	146,899	24,606	2,197	62.35	63.43	47.21
OmegaT	198,309	40,018	2,301	22,895	4,783	363	11.64	11.95	15.77	120,356	24,344	1,497	60.69	60.83	65.05

* Repetitions are part of the total of elements but not counted for elements with matching greater as 50%

Table 2: Empirical evaluation

5. CONCLUSION

As a conclusion, it can be said that in this paper it has been observed how these computer-aided translation tools greatly help the work of translators by speeding up and even reducing the translation process. Thanks to these tools, the work that could take months of work for the translator can be solved in a few days. This is why practically all translation companies, if not all, use this type of tools.

Throughout this paper multiple tools have been seen and compared and it has been observed that there is none that surpasses all aspects to others, therefore, it is already the task of the translator to decide which is the tool that best suits his taste and needs and which is the one that offers him the most benefits with respect to others.

Last but not least, it is important to note that, as its name suggests, these applications assist the translator, at no time they replace him. It is the task of the translator to provide and grow the translation memories that these programs use to develop their work and, even in the hypothetical case in which the tool had a translation memory so good that it would allow it to develop a perfect pretranslation, the task of reviewing and post-editing the document, if there was anything to change or improve, would still be in the translator's hands.

6. BIBLIOGRAPHY

6.1. Works cited

Berglund Prytz, Y. (2012). Types of Corpora and Some Famous (English) Examples. Retrieved from https://weblearn.ox.ac.uk/access/content/group/3a217dfd-a8cd-4034-8564-c27a58f89b9b/Handouts/CorpusTypes.pdf

Biber, D. (1995). *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

Carl, M., & Way, A. (2003). *Recent Advances in Example-based Machine Translation*. Dordrecht: Kluwer Academic Publishers.

Corpas Pastor, G., & Seghiri Domínguez, M. (2010). Size Matters: A Quantitative Approach to Corpus Representativeness. In R. Rabadán (Ed.), *Lengua, Traducción, Recepción. En Honor de Julio César Santoyo/ Language, Translation, Reception. To Honor Julio César Santoyo*. León: Universidad.

De Haan, P. (1992). The Optimum Corpus Sample Size?, in G. Leitner (Ed.), *New Directions in English Language Corpora: Methodology, Results, Software Development*. Berlin — New York: Mouton de Gruyter.

Faya Ornia, G. (1996). Revisión y Propuesta de Clasificación de Corpus. *Babel Revue Internationale De La Traduction / International Journal of Translation*, 60(2), 234-252. doi: 10.1075/babel.60.2.06fay

Fernández-Rodríguez, M. (2010). Evolución de la Traducción Asistida por Ordenador. de las Herramientas de Apoyo a las Memorias de Traducción. *Sendebar, 21*, 201-230. doi:10.30827/sdb.v21i0.374

Forcada, M. (2017). Making Sense of Neural Machine Translation. *Translation Spaces. A Multidisciplinary, Multimedia, And Multilingual Journal of Translation*, 6(2), 291-309. doi: 10.1075/ts.6.2.06for

Heaps, H. (1978). *Information Retrieval*. New York: Academic Press.

Hutchins, W., & Somers, H. (1992). *An Introduction to Machine Translation*. London: Academic.

Lagoudaki, E. (2006). Wayback Machine. Retrieved from https://web.archive.org/web/20070325114619/http://www3.imperial.ac.uk/portal/pls/portallive/docs/1/7307707.PDF

Leech, G. (1991) "The State of the Art in Corpus Linguistics", in Aijmer K. and Altenberg B. (eds.) English Corpus Linguistics: Studies in Honour of Jan Svartvik. London: Longman.

McEnery, T., & Hardie, A. (2014). *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press.

McEnery, T., & Hardie, A. (2012) "Support Website for Corpus Linguistics: Method, Theory and Practice". Retrieved from http://corpora.lancs.ac.uk/clmtp .

Quah, C. (2006). *Translation and Technology*. Basingstoke [Angleterre]: Palgrave Macmillan.

Seghiri, M. (2014). Too Big or Not Too Big: Establishing the Minimum Size for a Legal Ad Hoc Corpus. *HERMES - Journal Of Language And Communication In Business*, 27(53), 85. doi: 10.7146/hjlcb.v27i53.20981

Tabor, J. (2013). CAT Tool Use by Translators: What Are They Using?. Retrieved from https://prozcomblog.com/2013/03/28/cat-tool-use-by-translators-what-are-they-using/

Zanettin, F. (2002). Corpora in Translation Practice. Retrieved from https://www.researchgate.net/publication/228806527_Corpora_in_translation_practice Web. 5 Feb 2019

The History of SDL's Translation Software. (2019). Retrieved from https://www.sdltrados.com/about/history.html

What Is Machine Translation?. (2019). Retrieved from https://www.sdltrados.com/solutions/machine-translation/

6.2. Resources

Atril. (2019). Déjà Vu X3 (Version 9.0.765) [Windows].

Lexitrad. (2019). ReCor (Version 2.0) [Windows]. Universidad de Málaga.

MemoQ. (2019). MemoQ (Version 8.7.11) [Windows].

OmegaT. (2019). OmegaT (Version 3.6.0) [Windows].

SDL Trados. (2019). SDL Trados Studio 2019 (Version 15.0.0.20974) [Windows].

Wordfast. (2019). Wordfast Pro 5 (Version 5.6.0) [Windows].