



Universidad de Valladolid

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

Grado en Traducción e Interpretación

TRABAJO FIN DE GRADO

**FREE CAT TOOLS AS AN ALTERNATIVE TO
COMMERCIAL SOFTWARE: OmegaT**

Presentado por Veronica Nicoleta Anica

Tutelado por Ana María Alconchel

Soria, 2014

Content

ACKNOWLEDGEMENT	4
I. INTRODUCTION	6
1. Connection with competencies	7
1.1. General competencies.....	7
1.2. Specific competencies.....	8
1. PURPOSE	10
2. METHODOLOGY	11
II. THEORETICAL APPROACH	13
1. Translation and technology	13
1.1. Technological advances and the process of globalization.....	13
2. Benefits of technology in translation	15
3. New tendencies in translation.....	18
4. What are CAT tools.....	20
4.1. Translation Memory (TM)	22
4.2. Terminology-Management Systems.....	24
4.3. Alignment tools	26
4.4. Localization tools	28
5. Free and open-source software.....	33
III. ANALYSIS.....	38
1. Introduction to analysis.....	38
2. OmegaT - integrated translation environment	39
2.1. Project properties.....	39
2.2. Editing field behavior	41
2.3. Translation memory	43
2.3.1. Manipulating TMs	46
2.4. Terminology management.....	47
2.4.1. Dictionaries	47

Free CAT tools as an alternative to commercial software: OmegaT

2.4.2. Glossaries	47
2.4.3. Spell checker	49
2.5. Machine translation.....	51
2.6. Automatic Java Properties Aligner	52
2.7. Plugins	52
2.7.1. Tokenizer plugin.....	52
2.7.2. LanguageTool plugin	52
2.7.3. Scripting plugin	53
3. Impressions	55
4. Translating with OmegaT	58
4.1. OmegaT results	62
5. Advantages and disadvantages of OmegaT	66
IV. CONCLUSIONS	70
BIBLIOGRAPHY.....	72
ANNEXES	74

ACKNOWLEDGEMENT

Firstly, I would like to extend my deepest gratitude to Miss Isabel Comas Martínez for providing me help and guidance, generous support and for impelling my interest in CAT tools, reason for which my next academic step is heading towards this direction. I would also like to thank Dr. Judith Carrera Fernández for her valuable time and helpful comments.

Finally, very, very special thanks belongs to Răzvan Dumitrescu for his support and endless patience.

RESUMEN: El presente trabajo fin de grado trata sobre las herramientas gratuitas de traducción asistida por ordenador, en concreto, las características de OmegaT. La parte teórica del trabajo se centra en el impacto que tienen la tecnología y la globalización en el ámbito de la traducción. Consecuentemente, el trabajo aporta una aproximación teórica elemental a las diferentes herramientas TAO disponibles en el mercado. El trabajo intenta explicar los sistemas tecnológicos específicos más prevaletes (memorias de traducción, sistemas de gestión terminológica y herramientas de alineamiento y localización) para un mayor entendimiento del funcionamiento y de la utilidad de las herramientas TAO gratuitas y, consiguientemente, para reconocer la importancia de introducirlas como asignatura en el plan de estudios de Traducción e Interpretación. Por último, el trabajo pretende aclarar el concepto de software gratuito y de código abierto, puesto que la segunda parte del trabajo examina una herramienta de este tipo. La segunda parte del trabajo se dedica al análisis de OmegaT, una herramienta gratuita y de código abierto. Por último, se mostrarán los resultados obtenidos después de haber traducido una página web con esta herramienta.

Palabras clave: herramientas de traducción asistida por ordenador, memorias de traducción, sistemas de gestión terminológica, OmegaT, software gratuito y de código abierto.

ABSTRACT: This bachelor thesis deals with free computer-assisted translation tools, specifically OmegaT features analysis.

The theoretical part of the bachelor thesis concerns about the impact of technology and globalization on translation environment. Subsequently, the paper intends to provide an elemental theoretical approach on the different CAT tools available on the market. In order to be able to acknowledge the functioning and usefulness of free CAT tools and also the importance of implementing them as a subject in early Translation and Interpreting degree, the thesis illustrates the most prevalent specific technology such as translation memories, terminology-management systems, alignment, and localization tools. Lastly, the paper clarifies the concept of free and open-source software, as the second part of the paper is dedicated to the analysis of a tool like this, OmegaT. Ultimately, the results obtained after translating a webpage with OmegaT will be shown.

Keywords: computer-assisted translation tools, translation memories, terminology-management systems, OmegaT, free and open-source software.

I. INTRODUCTION

We have chosen to talk about free CAT tools because we consider that they are very helpful and valuable for both translators and translators to be, as they increase the productivity of the translation process.

Not long ago, the pen and paper were considered translators' tools, however, today the term *tool* has acquired new connotations. The development of technology has forced the translator to change his *modus operandi*, to adapt to new requirements, to surpass themselves in order to handle changes. If the incorporation of CAT tools in translator's everyday practice was once a challenge, at present time, they are indispensable, which is why CAT tools market is an exclusive one.

Fortunately, soon, free computer-aided translation tools have arrived and translators have been given the chance to continue using them without being conditioned by costs.

As there is a common tendency to believe that free goods are not as qualitative as priced ones, at the beginning, translators might have been reticent in using free CAT tools. With this paper, we are trying to explain and prove that free CAT tools can be as efficient as commercial ones.

Furthermore, we strongly believe that not only professional translators should benefit from CAT tools, but also translation students. Free CAT tools have improved considerably and they could be used in the classroom to help students learn how to use them, as they can continue using them when becoming professional. Even if they will choose to become freelance translators or to work in a translation agency, free CAT tools could be used without *constraints*, as a universal format has been created.

We are making these statements with certainty because we have studied specific subjects such as CAT tools, Localization, and Terminology management and we had the chance to employ free ones as well.

1. Connection with competencies

In this thesis, we find reflected general and specific skills that we have acquired along our Translation and Interpreting degree. We were able to formulate statements, to give arguments and to draw conclusions based on our reflections and observations on a particular free and open-source CAT tool. Also, we were able to show how this tool helps its users in the different translation phases.

1.1. General competencies

The general competencies appointed in the teaching guide of the bachelor thesis derive directly from the Royal Decree 1393/2007 of 29 October, the Law 3/2007 on Gender Equality Act, the Law 51/2003 on Non-discrimination and accessibility for persons with disabilities and also the Law 27/2005 on the Promotion of Education and the Peace Culture.

Therefore, the general competencies included are:

- G1. That students demonstrate knowledge and understanding in the area of study (Translation and Interpretation), which stem from the basis of general secondary education and often found at a level that, while it is supported by advanced text books, it also includes some aspects that involve knowledge from the forefront of their field of study.
- G2. That students know how to apply their knowledge to their work or vocation in a professional manner. Also, they have to have the competencies that are usually evinced through preparation and defense of arguments, and problem resolution within their area of study (Translation and Interpretation).
- G3. That students have the capacity to collect and interpret relevant data (normally within their area of study) to make judgments that include a reflection on key issues such as social, scientific or ethical.
- G4. That students are able to convey information, ideas, problems, and solutions to specialized and non-specialized audiences.
- G5. That students develop those competences needed to undertake further studies with a high degree of autonomy.
- G6. That students develop an ethical commitment in their development as professionals, a commitment that should enhance the idea of integral education, with critical and responsible attitudes, ensuring effective equality of men and women, equal opportunities, the universal accessibility of people with disabilities and the inherent values of a culture of peace and democratic values.

1.2. Specific competencies

As regards to specific competences, in this thesis we also identify specific competences related to the topic of our paper. Also, we were able to show how the tool we have chosen to analyze helps in the different translation phases. Moreover, we applied our acquired specific skills in defining CAT tools, categorize them, and show their main functions and components: translation memories (TM), alignment tools, terminology management, and localization. Lastly, we were able to extract conceptual information from the OmegaT analysis to emphasize on the quality that this tool provides.

The following competences are the most conspicuous in this thesis:

- E17. Mastering basic concepts on the management of computer tools that facilitate both their use and their integration in the work of the translator.
- E18. Using basic computer tools as specific apparatus for assistance in the different phases of the translation process.
- E20. Learning about the most appropriate localization tools.

Effectiveness, diversity of features that are intended to enhance the translation process, ease of use, different supported formats and files are just some of the advantages of using CAT tools. More and more free CAT tools are designed today to facilitate the translation of digital content and software, as localization is indispensable in an evolving technological society.

- E26. Being acquainted with the importance of the technology input in the translation process.
- E27. Applying the knowledge of computer tools for managing texts for general/specialized translation projects.

Subjects such as Documentation for Translators, Computing for Translation, Terminology, Terminology Management, Computer-Assisted Translation, Localization, ICT's for Translation have risen our awareness on the relevance and usefulness of the technology in the translation process introducing us different tools, resources, and projects such as [DocuTradSo](#).

Through advances in technology, we can perform many translation related tasks, reason for which is it utterly necessary to acquire these above mentioned competencies. As technology is evolving, computer-assisted translation tools are rapidly developing as well and, to increase productivity, it is important to be aware of their existence, to know how to use them properly as they represent a significant asset for the translation community.

- E28. Being acquainted with the aspects related to the task of general/specialized translation and the various processes involved in it.

Free CAT tools as an alternative to commercial software: OmegaT

- E34. Applying to general/specialized translation projects professional computer-assisted translation tools (CAT) and terminology management tools.

We were able to develop these competences by applying the knowledge and the skills we have acquired in technology-related subjects to specialized translation projects in subjects in which we were simulating real translation projects: in small groups, by rotation, we were terminologists, project managers, translators, and revisers.

The development in technology has consequences on translation in all its domains: legal and administrative, technical, literary, scientific, commercial or institutional translation. This thesis demonstrate that the use of translation tools facilitate both general and specialized translation process, as it is conveniently to search and manage translation projects and terminology databases.

Moreover, we applied our acquired specific skills in defining CAT tools, categorize them, and show their main functions and components: translation memories (TM), alignment tools, terminology management, and localization.

Lastly, we were able to extract conceptual information from the OmegaT analysis to emphasize on the quality that this tool provides.

- E40. Designing and developing dictionaries and terminological databases.

To conclude, with this paper we aim to prove that we have acquired the required competences to carry out this type of investigation project:

- E47. Showing management skills and assessment of the quality of the compiled information skills, which will serve as empirical support for a research project.
- E51. Learn about the interdisciplinary foundations that will serve as a theoretical framework for the bachelor thesis.
- E52. Ensure the quality of the thesis within specific deadlines.

Furthermore, we will explain the purpose of this paper and its aim, as regards to the competencies mentioned above.

1. Purpose

This paper is, as the title suggests, about free computer-aided translation (CAT) tools. The reason why we have chosen this topic is because we have apperceived that, although computer-aided translation tools are and have been the subject of many studies and discussions, not all translators are familiarized with the concept.

The aim of this paper is to explain and exemplify the imperative necessity of acknowledging the presence of free computer-aided translation tools in the information era as they increase translators' efficiency and productivity.

It is no longer a question whether to employ or not a CAT tool, as the fact that computer-aided translation technology is helping translators in being more efficient is a reality already assumed and greeted. The main purpose of this thesis is to inform about the applicability of the free CAT tools available on the market by analyzing the latest version of OmegaT 2.6.3, a free and open source CAT tool, in order to stimulate interest and direct efforts towards an extended use of these kind of tools.

Learning about the dimension, functioning, and management of free CAT tools is the best long-term investment because the translator to be will manage to dominate the tool properly both in class and real life assignments. Furthermore, we consider that the different CAT tools, especially free CAT tools, and their applicability should be a compulsory subject in Translation and Interpreting syllabus.

2. Methodology

In this chapter we will explain, among other matters, the framework of this paper, the reason why we have chosen to analyze OmegaT, the phases of this analysis, the criteria we have used to analyze the tool and the motives of choosing this criteria.

The methodology we have used in order to understand better how these free computer-aided translation tools function and the benefits they bring when employing them in translation practice can be divided in four main parts: the introduction, the theoretical approach, the analysis of a particular CAT tool, and the conclusions.

In the first part we will see the importance of the subject matter, its connection with the competencies acquired in the Translation and Interpreting degree, the purpose of this paper and its methodology.

In the second part of this thesis we will explain the relation between translation and technology and the advantageous results of this fusion for the translation community. Also, we will examine the impact of technology and globalization on translation practice to acknowledge to which extent it is imperative to introduce CAT tools not only in professional working context, but also in university syllabus. We will go further and explain new tendencies in translation which are being caused by this relation in order to highlight the importance and the impact of free CAT tools. Furthermore, we will define and categorize CAT tools, and we will go through their main functions and components: translation memories (TM), alignment tools, terminology management, and localization. In addition, we will explain the concept of free and open-source software and we will explain why is it important.

In the third part of the paper we will analyze a stand-alone free and open-source CAT tool: OmegaT 2.6.3. A stand-alone tool or an integrated translation environment refers to the fact that the program has incorporated the translation memory and the terminology-management systems, the editor in which to perform the translation and other additional features. The advantage of such tool is that we have everything in one place ensuring greater stability and customization. This is one of the main reasons we have chosen to test OmegaT.

At the beginning of this paper we have explored many CAT tools searching for those which are free and open-source, but some of them were free and not open-source, others were open-source but not free and many of them were not integrated environments:

- *Anaphraseus* is a xliff editor, an OpenOffice macro which works as a integrated document, not environment.
- *Heartsome* is a translation memory editor, therefore not suitable for our analysis.
- *Across* is free only for freelance translators and students and it is not open-source.

Free CAT tools as an alternative to commercial software: OmegaT

We also considered comparing OmegaT with *Virtual* or *ForeignDesk*, both free and open-source, but they were not as complete as OmegaT, therefore, it would not have been representative for our analysis.

In the third part we will analyze OmegaT, we will examine the interface and its functioning, and we will go through all its features: project properties, translation memory, terminology management, machine translation, aligner, and plugins. We will focus on providing insights of this tool to see if its users can obtain very good results as in quality, productivity, and efficiency with a free and open-source CAT tool. We have chosen this analysis criteria because we considered it was the most appropriate one as these are the main features of the tool.

Furthermore, we will state our opinion about this tool and we will perform an exercise translating a webpage with OmegaT to observe its running when translating other format than .docx. Subsequently, we will show the results of the exercise along with advantages and disadvantages of OmegaT that we will apperceive.

In the fourth and last part of the paper we will draw our conclusions and we will add the bibliography we have used for this thesis.

II. THEORETICAL APPROACH

1. Translation and technology

Trying to define translation is a difficult task, which many language experts of all times have tried to do. We strongly believe it is not fair to affirm that one definition is more accurate than another, as the definition of translation is as flexible and subtle as the process itself. How can we say that Nida & Taber's definition or Newmarks' were inaccurate? Of course, we are in no position to pronounce a verdict, but we can state the following: nowadays linguists, translators and language experts find these definitions incomplete, but only as in relation to the time and place we find ourselves in. Therefore, we simply proclaim that the process of translation can be defined only taking into account time, space, culture, society or a specific group of society.

This is the reason for which we consider this definition the most appropriate one: "Translation may be defined as follows: the replacement of textual material in one language (SL) by equivalent textual material in another language (TL). This definition is intentionally wide - not vague, though it may appear so at first sight." (Catford, 1965: 20)

As to present time, experts try to formulate an empirical definition that can be veracious no matter the time, place, and context it relates to. However, the first thing to say about this definition, that we know with certainty and, in fact, is the matter that concerns our present paper, is that it will be undoubtedly connected with technology.

1.1. Technological advances and the process of globalization

Technological advances are perceived as the main determinant in the process of globalization. We usually associate technology with a set of tools designed for helping us either at work or on our everyday tasks, but the truth is that technology is a powerful weapon that has inordinately changed our world. It encloses a series of elements that have contributed to globalization, which implies worldwide integration of social, political, business, and cultural systems. (Lawlor, 2007:1)

Telecommunications, computer hardware and software, IT applications have increased our ability to use and share information for personal, political, and commercial purposes. The Internet and the World Wide Web have helped us communicate, collaborate and stay up to date. The history books will refer to the modern world as a global village connected by technology. Improved information technology has been the main facilitator and driving force of globalization on all environments. (Lawlor, 2007:1)

There are many advantages that the IT revolution and the globalization have brought, for example, the possibility of finding jobs, services, and cheaper products with one click, the

development of the e-commerce and therefore increased competition, and the innovations in all professional areas, translation included. Without any doubt, globalization's repercussion on business environment is a highly disputed topic, but for this paper, we will aim our attention at the translation industry because there is a growing demand for translation as English is not the *lingua franca* of the global market economy or of the world of technology and "the evidence is in your computer, in Word, go to Tools/Language/Set Language [...]." (Pym, 2004: 2)

Therefore, it is the translator's job to connect people, to unite them through linguistic similarities and differences by creating bridges using world's most valuable tool: words. Nonetheless, in the next chapter we will further explore this perspective.

2. Benefits of technology in translation

Lynne Bowker (2002) beautifully manages to express the relevance of technology in translation business citing Samuelsson-Brown (1996, 280):

Technology is developing at a frightening pace and the demands made on the translator do not show any signs of abating. In fact, the translator is becoming more and more dependent on information technology and, if the translator does not adapt to change, he or she may become uncompetitive. (Bowker, 2002: 3)

The impressive development in technology has direct repercussions on translation in all its realms: legal and administrative, technical, literary, scientific, commercial or institutional translation. All these effects have everything to do with *productivity* and *efficiency*, attributes that are more than necessary taking into consideration that we live in a society where time is measured in 0 and 1 binary digits which generates data at the speed of light. Now translators have constant access to Internet, therefore to information, they can consult parallel texts for better understanding the source text they are dealing with, thousands of online resources like dictionaries, web forums, mailing lists, usenet newsgroups, etc. or even interact in real time with other translators. If all the above mentioned resources are not an option, translators can digitalize most of the data they work with so that they can handle it easily, that is, pre-edit, edit, and post-edit, as many times as they wish.

It is important to point out that translators are users of technology rather than developers (although this is not excluded as we will see further on), but the outgrowth of this astute fusion, translation and technology, has changed the translation business radically contributing to its development.

As mentioned before, *efficiency* and *productivity* are increased because translators have the possibility to create corpora, terminology-management systems (TMS), use translation memories (TM) and bilingual concordances, create their own glossaries, or translation projects. In other words, they have now integrated the computer-aided translation (CAT) tools in their professional translation workplace, which have sped up the translation process.

Many translators have incorporated CAT tools in their workstation because of there are many advantages of using CAT tools from which the most significant ones are TMs and TMSs.

Translation memory systems facilitate the translation and ensure coherence across translation. TMs are one of the best innovations of the modern world for translators, because "No other technology has changed the general conditions of translation as a professional service as radically as TM systems have done over the past 20 years." (Reinke, 2013: 1)

Free CAT tools as an alternative to commercial software: OmegaT

Terminology-management systems provide quality as they ensure consistency within a translation project because we can store specific terminology in termbases. Without any doubt, this speeds up the process of translation as we can search through our own files or simply we do not need to perform the same terminology search twice because some TMS have the automatic terminology lookup feature. Furthermore, besides flexible storage and fast retrieval, we have the possibility to share a termbase with our client or with other translators, whether for consulting or contributing to that termbase. Nowadays, TMS's format is no longer an impediment as there are many programs we can use for converting if the TMS we are using cannot handle the universal format known as Term Base eXchange (TBS). (Somers, 2003: 4)

Several studies reveal that using a CAT tool could result very beneficial for translators because CAT tools are designed for facilitating the translation process, increasing translators' efficiency and productivity considerably. Fotini Vallianatou, a freelance translator with a degree in Translation from the Ionian University in Greece, has conducted an exercise calculating her productivity over a period of 18 months using different CAT tools. She states that the results were impressive and, undoubtedly, both her efficiency and productivity increased keeping in mind that she translated from English into Greek technical material, mainly automotive and medical instrumentation. Therefore, the productivity depends "on the type of job and the memory and terminology databases used [...] the more specific the memory databases, the larger the word volume you can translate in the same time period." (Vallianatou, 2004: 4)

In consequence, we can say that especially when working with texts from a specific field, that contains repetitive terminology and phraseology, CAT tools are recommended because the TM and the TMS will provide good translation suggestions.

There is not one correct answer to the question *which CAT tool should we use?* because translators choose their tools following different criteria depending on their needs: client compatibility, operating system compatibility, price, license type, etc. According to Reinke (2013), the most well-known commercial tools are Across, Déjà Vu, memoQ, MultiTrans, SDL Trados, Similis, Transit and Wordfast. Nevertheless, if we are to ask translators on business-related social networking sites for translation such as [ProZ](#), they all agree on a list of the most common CAT tools among which we can find OmegaT. TranslateMedia, a digital and technology company that has developed its own translation management technology as it is working with professional linguists and translators for its digital agencies, has created a list of the most common and best known CAT tools according to ProZ (see Annex 1). The reliability of ProZ may be disputable, it might not be 100% reliable source, but it is, nonetheless, a worldwide community of language professionals and we have chosen this list because it is highly illustrative. (TranslateMedia, 2014)

All in all, the impact of technology on translation business is a very positive one: the digitalization of the data, the constant access to Internet and the integration of CAT tools in

Free CAT tools as an alternative to commercial software: OmegaT

translators' professional routine as many "users in industry and international organizations usually claim a 25 to 60 per cent rise in productivity." (Reinke, 2013: 2)

3. New tendencies in translation

There is still *on the table* the argument concerning the translators' visibility, but we have to admit that the technology boom and the internet revolution have started to give translators more and more recognition for their significant addition to nowadays society. How could Internet content be free, open and accessible for all users regardless their language without translators' colossal contribution? English used to be the prevailing language on the Internet until recent years, but now user authorship is polyglot thanks to the outstanding translators' work of localization. This leads us to a new complex process that we would like to define in order to understand it better: localization.

Lyngo Systems (2002: 4) defines localization as "the process of customizing a product for consumers in a target market so that when they use it, they form the impression that a native of their own country designed it." Hence, the localization of a product implies larger audience and, therefore, more economic benefits. The accelerated development of technology had lead to a conspicuous increment of computer users and, therefore, to an upsurge of "variety of electronic documents formats" that diffuse over the Internet. Thence, translators were required to have ample computer science for "the translation of the documentation interfaces and help files included in computer software applications and the translation of websites". (Alcina, 2008: 1)

Given these facts, we can say that localization is closely linked with technology and globalization as localization is seen as "a process to facilitate globalization by addressing linguistic and cultural barriers" with the help of technology. (O'Hagan, 2002: 66) To this extent, translation and localization have progressively changed the translation practice because nowadays this offline process has transferred online and collaborative.

As regards to the collaborative capability of the Internet, it is interesting to mention the relatively recent creation of a new phenomenon: the crowdsourcing. This consists in the cooperative Internet task solving, which was initially posted by an individual, a company or, why not, an institution. Indubitably, companies are the one that gain more benefits from crowdsourcing, considering Heer and Bostok's (Heer & Bostock, 2010: 1) affirmation "Crowdsourcing is a relatively new phenomenon in which web workers complete one or more small tasks, often for micro-payments on the order of \$0.01 to \$0.10 per task" as they are seen as rewards, rather than payments.

Unfortunately, crowdsourcing has some facets that affect the professional translation market in terms of that many translation jobs are carried by amateurs, i.e. language connoisseurs, and not experts in the field, which implies more competition on the market and lower rates as well. Although the reasons for which these people are involved in crowdsourcing

are not financial, it is a common practice to reward them with money for their dedication and help.

While crowdsourcing is blurring the boundaries between amateurism and paid work, and between working and leisure time, increasingly powerful and well-performing technological tools – machine translation systems and other CAT tools – are appearing. All these changes are often frowned upon as jeopardizing the very survival of the translation profession. (European Union, 2012: 3)

Therefore, crowdsourced translation is a topic of great concern because, with the high demand for translation and free CAT tools, amateurs are threatening the translation business market.

The demand for translation and language services in general is exploding and the resources available cannot keep up with it. These new tools, together with the other changes and developments connected to the new Internet culture, therefore, appear as the only viable option to help translators cope with the pressure they are to work under. The new developments will not sweep away translators, but it is a fact that they will impose far-reaching adaptations in the way the profession is conceived and performed. (European Union, 2012: 3)

This is the natural evolution of the global market and we are not saying we should accept everything as it is as long as we have the power to change what might represent an impediment in the development of our jobs. Instead, what we strongly defend is that technology has brought to the translation industry a series of advantages, which we should all exploit for our benefit.

4. What are CAT tools

Before moving on to the core of our paper, which is to have a close look at free license CAT tools, we must explain first what are CAT tools and which are their characteristics, and also to define some related concepts so that we make sure everything is clear.

Lynne Bowker (2002: 4) in her book explains the difference between "human-assisted machine translation (HAMT), which is often shortened simply to machine translation (MT), and machine-assisted (or -aided) translation (MAT)", clarifying that nowadays we usually refer to these last ones as computer-aided (or -assisted) translation (CAT). Furthermore, she briefly, but precisely and unequivocally, describes the particularity of both MT and CAT:

The major distinction between MT and CAT lies with who is primarily responsible for the actual task of translation. In MT, the computer translates the text, though a human translator may later edit the machine output. In CAT, human translators are responsible for doing the translation, but they may make use of a variety of computerized tools to help them complete the task and increase their productivity. Therefore, whereas MT systems try to replace translators, CAT tools support translators by helping them to work more efficiently. Lynne Bowker (2002: 4)

Somers also stresses the importance of understanding the difference between CAT and HAMT:

[...] often a finer distinction is made between Machine-Aided Human Translation (MAHT) and Human-Aided Machine Translation (HAMT) implying a distinction between a basically human activity involving computer-based tools on the one hand, and a computer-driven activity requiring the assistance of a human operator. (Somers, 2003: 2)

Austermühl states that the main objective of CAT tools is to provide the translator with all the "information needed during the reception, transfer, and production phase of the translation process" and he describes them as follows:

[...] while accepting their advantages in assuring above all phraseological and terminological consistency and thus translation quality, we primarily characterize them as leverage or productivity tools aimed at an automated direct transfer of source text fragments into the target language, and as solutions for dealing with specific data formats of software-related texts. (Austermühl, 2001: 3)

In other words, CAT tools are, as the name itself states, a set of computerized means intended for helping the translators' work. Below we can see a table made by Lynne Bowker where we can see more easily the difference between the concepts we have described earlier.

HT	CAT	MT
<ul style="list-style-type: none"> • Word processors • Spelling and grammar checkers • Electronic resources (e.g., CD-ROMs) • Internet (e.g., WWW, e-mail) 	<ul style="list-style-type: none"> • Data-capture tools • Corpus-analysis tools • Terminology-management systems • Translation memories • Localization and Web-page translation tools • Diagnostic tools 	<ul style="list-style-type: none"> • Machine-translation systems

Figure 1. Lynne Bowker's classification of technology tools used in translation

We will concentrate only on the second column of the table, that is, CAT tools and, to be more precise, OmegaT, a free and open-source CAT tool.

There are several types of software in terms of license, price, and restriction of use and/or modification. We will briefly explain the concepts now, as later on we will have a close look at those that are closely related to our paper:

- Free or freeware software means we can download and use it free of charge for an unlimited period of time.
- Open-source software is a kind of software that allows us, if we have programming skills, to personalize it, to copy and distribute copies of the license software.
- Commercial software is a type of software designed for selling purposes. Usually, this kind of software is also shareware, which means we can use it free of charge from 15 up to 30 days, but after this trial period, we have to pay to continue using it.

What they all have in common is the fact that they work with almost the same algorithm, reason for which we will have a close look to a free CAT tool, as it does not differ too much from the commercial ones. Of course, the fact that the last ones enjoy financial benefits and a team whose purpose is to improve them constantly makes them popular among translators, but studies have shown that the algorithm used for these software has not change too much. (Bowker, 2002: 4)

Another reason for which many translators work with certain specific commercial software like SDL Trados or MemoQ is, in fact, the popularity they have gained in time. Most translation

agencies work with one or another and they recommend (and sometimes impose) to work with the same one. Not so many years ago, it was difficult for translators to defend their preferences in regards to CAT tools because they had to flex and adapt to the requirements of the translation agency they worked for or collaborated with, in the case of freelance translators. This problem was associated with the import or export process of the translation memory, a database that is used for storage both of the source text and its translation. However, developers designed a format called TMX (Translation Memory eXchange) which made the transfer of translation memories much easier because almost all CAT tools have integrated it, even the free ones.

Predominantly all CAT tools have integrated a series of specific technology such as translation memories (TM), terminology-management systems, alignment of the text source and the translated text, and localization tools. It is very important to mention that there are other CAT tools, but these ones are worth mentioning as they are the ones that bring changes that are more significant in conventional translation practice.

We will examine all these components individually in order to fathom how CAT tools work and, consequently, to be able to appreciate the quality and usefulness of a free CAT tool.

4.1. Translation Memory (TM)

Translation memory is a database in which the source text is saved with its correspondent translation. It is very accessible to work with as it divides the source text into segments where the user or the tool can insert the matching translation.

Segment	Source text (EN)	Target text (ES)
	For more information, ask your doctor or pharmacist.	Para más información, consulte a su médico o farmacéutico.

Segment	
Source text (EN)	For more information, ask your doctor or pharmacist.
Target text (ES)	Para más información, consulte a su médico o farmacéutico.

Figure 2. Example of how segments could be displayed

The segments, which represent sentences, can be displayed in two different ways, as we can see in Figure 2: aligned, the source text located on the left side and the translation on the right side, or as list. The display depends on the interface of the CAT tool we choose to work with.

The main purpose of a translation memory is to store the source text with its translation for future reuse. The most important convenience it has is the fact that all translations are kept in one place and they can be reiterated with promptness. Lynne Bowker defines this process of reusing previous translation as leveraging:

Reusing a previous translation in a new text is sometimes referred to as "leveraging." Although language is dynamic, it is quite repetitive, and people often use the same or similar expressions when communicating similar ideas. The volume of translation is increasing, and most translators have had the experience of being asked to translate a text containing passages that they (or they colleagues) have translated on a previous occasion. (Bowker, 2002: 93)

Harold Somers says that the idea of a TM is that "the translator can consult a database of previous translations, usually on a sentence-by-sentence basis, looking for anything similar enough to the current sentence to be translated, and can then use the retrieved example as a model". (Somers, 2003: 2)

As Somers, Austermühl refers to translation memory systems as databases:

[...] databases that allow for the parallel storage of source and target text segments on a sentence, phrase or word level. These segments can be retrieved and then used for a new translation project. Special software identifies whether the new text to be translated contains passages that are already stored in the translation memory (TM). The translations corresponding to the segments found in the TM database are then offered as translation proposals for the new text. (Austermühl, 2001: 3)

From all the definitions above mentioned we can deduce that, in essence, translation memories are very useful for texts that have repetitive terminology and phraseology such as legal and administrative, technical, scientific, commercial, institutional translation or medical. This is also stated by Alcina (2008), with the mention that CAT tools have great importance in localization:

Computer-assisted translation (essentially consisting of translation memories) and machine translation are useful in specialized translation, that is, when dealing with texts containing a large amount of terminology and repeated structures. Computer-assisted translation tools are also of great importance for localization purposes, in other words, for translating computer programs, video games and websites. (Alcina, 2008: 79)

In the case of these types of texts, the foremost used lexicon is almost the same and, with a solid TM, it can be retrieved quickly without the need of performing the same terminology

search every time a translator faces a text as such. For this reason, it has been proved that TM is not very efficient when it comes to literary translation, as it depends more on creativity.

Translation memory systems are very complex. The way they operate consists mostly in providing, when a new text is added or connected to the TM, exact, full, and fuzzy segment matches, and also term matches. We will briefly see what each one of them deems:

- Exact matches: when the new segment to be translated is identical from all points of view (punctuation, formatting, etc.) to a segment previously translated and saved in the TM, the TM will generate the retrieved translation.
- Fully matches: to some extent, this is a new type of matches that is still being developed, hence not very common for all CAT tools. It consists in the ability of generating the complete segment regardless of some subtle dissimilarities such as: numbers, dates, measurements, etc.
- Fuzzy matches: when the new segment to be translated is similar to a segment from the TM, this will generate the translation pointing out the differences to be corrected.
- Term matches: when none of the above matches is found, the TM offers specific terms translation. In addition, a TM can support integration of terminology management tool that gives terminology suggestions simultaneously with the translation process.

4.2. Terminology-Management Systems

Terminology management tools can be used as a separate tool or can be integrated into a translation memory. The main purpose of a TMS is to record terms into a term base so that the translator does not have to perform twice the same search. They are also very useful in maintaining coherence in translations, as the term base is usually specific for a field, i.e. technical or medical or for specific clients.

In relation to the time we use to spend on consulting external resources in the search for terminological equivalents, terminology-management systems represent some ideal tools taking into account the high demand and the pressure of the translation global market. They are used for storage, retrieval and editing of term entries. In addition, if we have a solid term base, we can export it as a glossary or as a dictionary, depending on the tool we use there are different options.

Previously, TMS used to have very basic functions, that is, they allowed the insertion of an equivalent pair in two languages and it could only be retrieved in one direction. To be more specific, if the term base was from source EN to target ES, then all the searches we have made should have been in English in order to get any results. If the search were made in the target

language, the term base would not have been able to recognize the search. However, today TMSs have developed considerably, and they allow us not only to choose multiple languages, but also to personalize the term entry so that we get results depending on the context we need that term. Below, we can see an example of a complete term entry that we used in a mycological terminology project at our university.

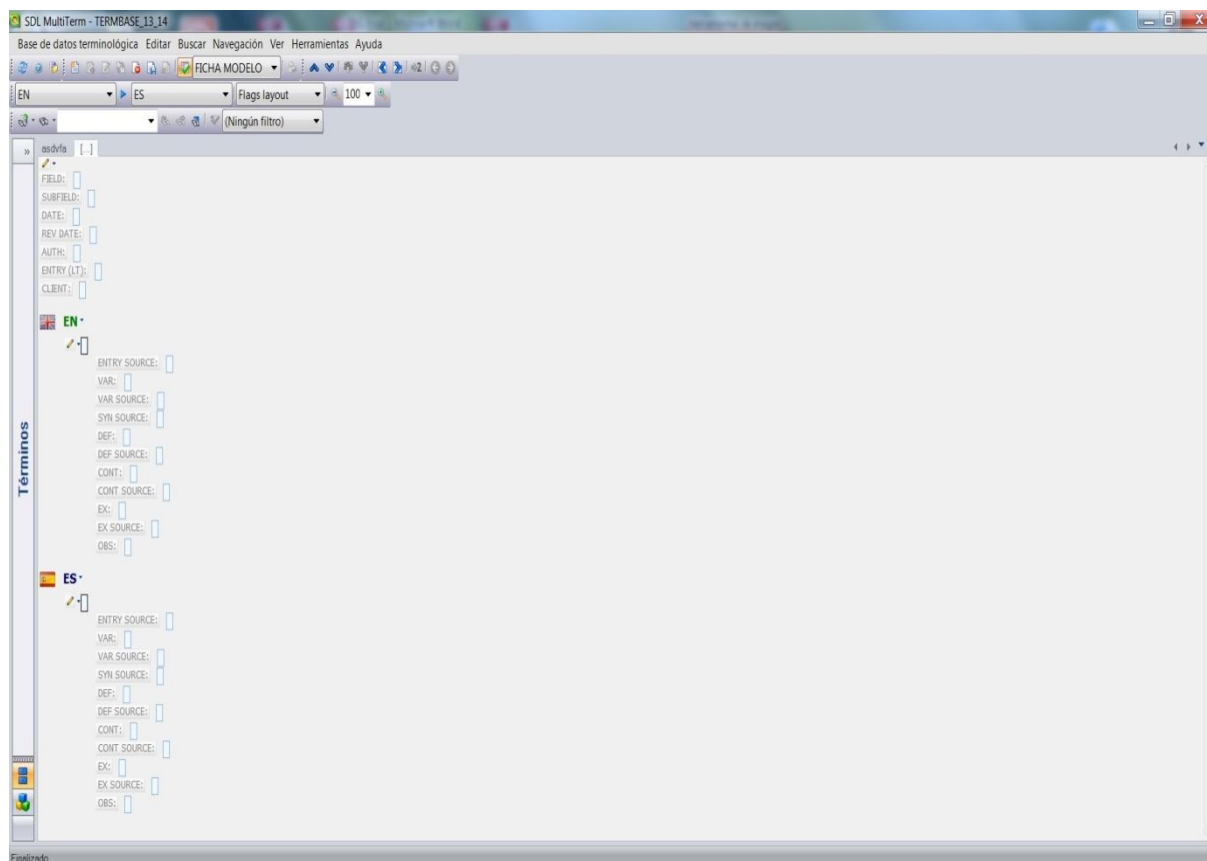


Figure 3. TMS MultiTerm entry model

The possibility of storing terms with a predefined field such as definition, context, variant, example, observation, etc. also makes easier the process of retrieval: for a specific client, author, reviser, field, subfield, date. It is interesting to mention that various term bases can be merged, reason for which the field *author* appears in Figure 2.

As well as translation memories, a TMS can be used as a stand-alone tool or it can be integrated into a TM. Unfortunately, exchanging term bases is not always an easy task as translators use different TMS tools that work with different formats. Nevertheless, experts have designed a universal interchangeable format, Term Base eXchange (TBX), but, as any other technological innovation, it could be enhanced.

Some TMSs will allow data to be exported directly to various word-processor or desktop-publishing formats or to be imported and exported according to international standards, such as the Machine-Readable Terminology Interchange

Format (MARTIF - ISO 12200). A new standard, known as Term Base eXchange (TBX), is currently under development by the Open Standards for Container/Content Allowing Reuse (OSCAR) special-interest group belonging to the Localization Industry Standards Association (LISA). (Somers, 2003: 59)

As regards to the file format, the OSCAR group has developed a standard format for translation memories and terminology-management systems, as mentioned above. These formats, both TMX and TBX, are revolutionary in the translation tool market because now translators can employ a single TM or TBX regardless of the one imposed by the client or the agency. All this, translated to time and money, it is, without any doubt, a huge step ahead considering the fact that the import and export process of the data between different TM and TMS is no longer a problem. Translators can choose the TM or TMS they prefer to work with, which means that a large number of CAT tools developers are losing money.

In relation to this direction, it is imperative to precise that many free CAT tools have incorporated these formats. This means that translators do not need to purchase a CAT tool they are not trained to use or they simply dislike because now the TMS data can be read by TMS systems, as well as TM data can be read by TM systems. As a matter of fact, this format related enhancement adopted by free CAT tools brings nothing but advantages and benefits to the translation community.

4.3. Alignment tools

Alignment is the process that allows the user to recycle translated content by connecting text, paragraphs, sentences or words with their translated equivalent. One of the most popular uses of these tools is sentence alignment. If we have important content that needs to be added to a TM, for example, this tool is ideal because we do not have to retranslate the text, but to use an alignment tool that examines the source text and the target text and links the corresponding sentences. Somers (2003) defines alignment as "matching up the source text and the translation segment by segment into translation pairs. [...] If the translation is straightforward, then so is the alignment."

However, this is an automatic process and, as all automatic processes, it cannot be perfect, reason for which is required a post-revision of the text to make sure that each source sentence is paired with the right target one. If the texts are differently structured, the tool is not able to identify these subtle variations such as:

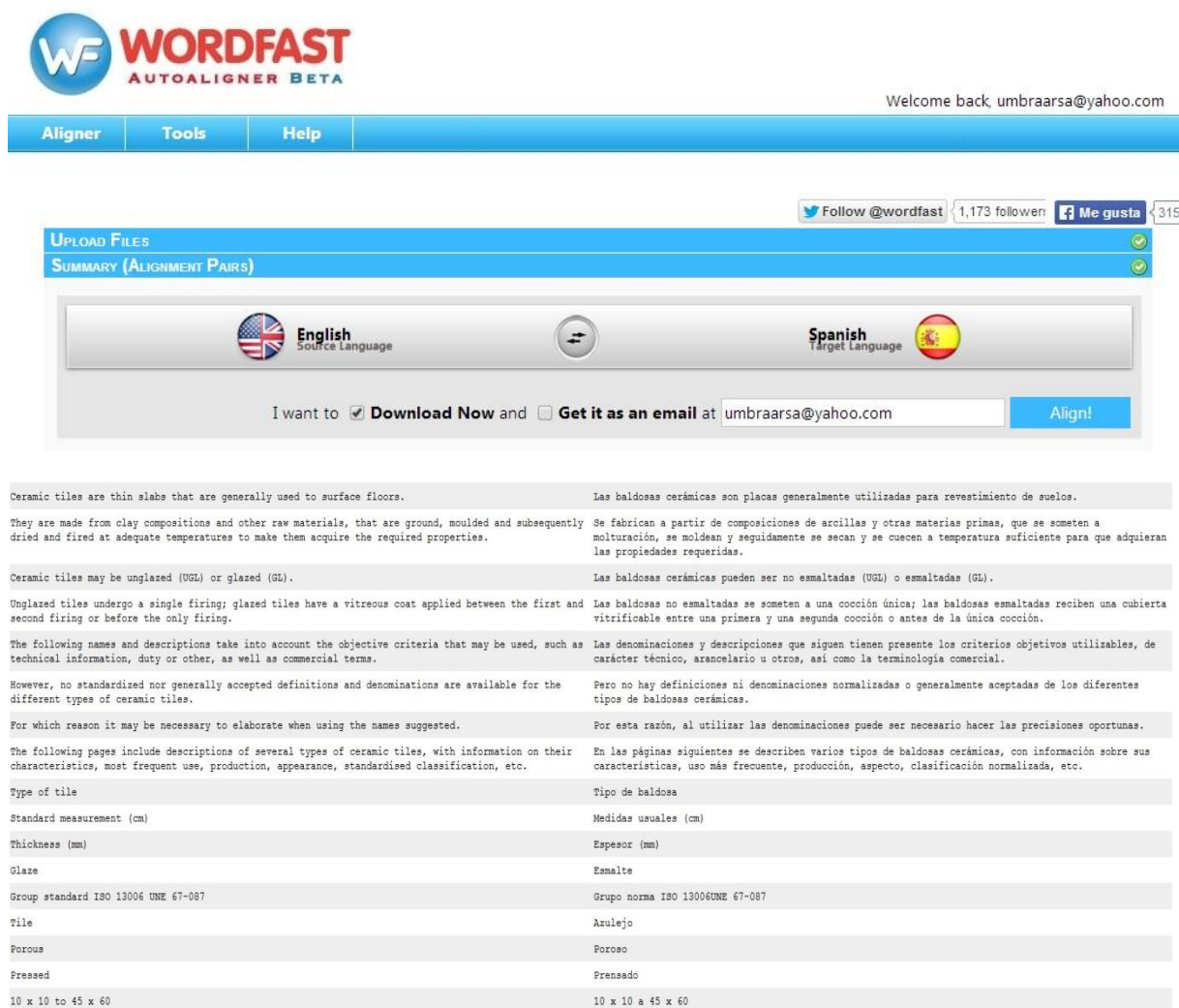
- Abbreviation or decimal points that could be misinterpreted as periods.
- The translation of one source sentence with two or more target sentences.
- Different word order in the translated text which could not be recognized and paired with the source text.

Free CAT tools as an alternative to commercial software: OmegaT

Therefore, the automatic alignment has its limitations, but with a simple intervention of the translator at the end of the process to fix these expected miscalculations, it becomes a powerful computer-aided tool.

This is excellent when it comes to build up a strong translation memory, but it is important to know that a TM created from scratch would give more accurate results than a TM created with aligned texts. As mentioned before, TMs work with segments, hence when a segment is translated and saved into the TM, this one will process it and use it for future similar or identical source text, and also to suggest translation for segments that are not very similar, process called propagation. This is why TMs that contain post-translation alignment cannot reuse internal repetitions.

Another interesting feature of the alignment tool is the file format that it supports. If we have previously translated content in other format than .doc, which is the most common one, we can still use the tool because many of them support nowadays a wide range of formats: HTML, STF files, Microsoft PowerPoint, Microsoft Excel, InDesign, TXT, XML, etc.



WF **WORDFAST**
AUTOALIGNER BETA

Welcome back, umbraarsa@yahoo.com

Aligner Tools Help

Follow @wordfast 1,173 followers Me gusta 315

UPLOAD FILES

SUMMARY (ALIGNMENT PAIRS)

English Source Language

Spanish Target Language

I want to Download Now and Get it as an email at umbraarsa@yahoo.com **Align!**

Ceramic tiles are thin slabs that are generally used to surface floors.	Las baldosas cerámicas son placas generalmente utilizadas para revestimiento de suelos.
They are made from clay compositions and other raw materials, that are ground, moulded and subsequently dried and fired at adequate temperatures to make them acquire the required properties.	Se fabrican a partir de composiciones de arcillas y otras materias primas, que se someten a molidura, se moldean y seguidamente se secan y se cocen a temperatura suficiente para que adquieran las propiedades requeridas.
Ceramic tiles may be unglazed (UGL) or glazed (GL).	Las baldosas cerámicas pueden ser no esmaltadas (UGL) o esmaltadas (GL).
Unglazed tiles undergo a single firing; glazed tiles have a vitreous coat applied between the first and second firing or before the only firing.	Las baldosas no esmaltadas se someten a una cocción única; las baldosas esmaltadas reciben una cubierta vitrificable entre una primera y una segunda cocción o antes de la única cocción.
The following names and descriptions take into account the objective criteria that may be used, such as technical information, duty or other, as well as commercial terms.	Las denominaciones y descripciones que siguen tienen presente los criterios objetivos utilizables, de carácter técnico, arancelario u otros, así como la terminología comercial.
However, no standardized nor generally accepted definitions and denominations are available for the different types of ceramic tiles.	Pero no hay definiciones ni denominaciones normalizadas o generalmente aceptadas de los diferentes tipos de baldosas cerámicas.
For which reason it may be necessary to elaborate when using the names suggested.	Por esta razón, al utilizar las denominaciones puede ser necesario hacer las precisiones oportunas.
The following pages include descriptions of several types of ceramic tiles, with information on their characteristics, most frequent use, production, appearance, standardised classification, etc.	En las páginas siguientes se describen varios tipos de baldosas cerámicas, con información sobre sus características, uso más frecuente, producción, aspecto, clasificación normalizada, etc.
Type of tile	Tipo de baldosa
Standard measurement (cm)	Medidas usuales (cm)
Thickness (mm)	Espesor (mm)
Glaze	Esmalte
Group standard ISO 13006 UNE 67-087	Grupo norma ISO 13006UNE 67-087
Tile	Azulejo
Porous	Poroso
Pressed	Prensado
10 x 10 to 45 x 60	10 x 10 a 45 x 60

Figure 4. Example of a free cloud autoaligner tool called WORDFAST Beta

There are also free alignment tools that perform the basic functions of an alignment tool. It is true that they have some limitations comparing to the commercial ones in terms of number of language pairs they can support or the number of files they align at a time, or format, for example. In Figure 5 we have an example of a free autoaligner tool, which it happens also to be in the cloud, that is, it is hosted on a webpage, so it works online without having to install it on our computer. It can allow the alignment of three source documents with their target texts simultaneously.

4.4. Localization tools

We have chosen the following quote because it shows how important is for translators to adapt to the new challenges of a continuously changing world. Chiew Kin explains the complexity of the process of localization:

Until the early 1990s, the time when the Internet began to be used worldwide, the translation types given in Hutchins and Somers (1992) were certainly applicable. More than a decade later, the boundaries of these four translation types have become more blurred. Although many writers in the field still make clear distinctions, these have become harder to maintain as technology becomes increasingly multifunctional and more multitasking. The pace of change in the development of translation technology is extremely rapid; what is current today may become outdated tomorrow. (Chiew Kin, 2006: 19)

As regards to localization, as we explained before, it is more than source-to-text linguistic transfer because it consists in a set of other processes. Acknowledging Chew's broad definition of localization being "the process of changing the documentation of a product, a product itself or the delivery of services so that they are appropriate and acceptable to the target society and culture" as well as the contracted one of Pym localization being "the adaptation and translation of a text (like a software program) to suit a particular reception situation", we will try to provide a more complete one from Schäler:

[Localization is] the linguistic and cultural adaptation of digital content to the requirements and locale of a foreign market, and the provision of services and technologies for the management of multilingualism across the digital global information flow. (Schäler, 2010: 209)

In other words, localization is more than source-to-text linguistic transfer, reason for which is often referred as in relation to globalization, internalization and translation.

Globalisation, Internationalisation, Localization, and Translation (GILT) are terms which nowadays are used very often. The borders between them are often

blurred, as many people consider localization as counterpart of globalization on the grounds that the former is "supposedly" based on a local and the latter on a global audience or market. (Anastasiou & Schäler, 2010: 11)

These concepts are related because localization has its roots in technology and, as technology is the present constancy of our now digital society, geographical references have become much closer, hence the need of content adapted for a specific culture.

To understand better what these concepts mean, we will enumerate their definitions proposed by the Localisation Industry Standards Association (LISA):

- *Globalization addresses the business issues associated with taking a product global. In the globalization of high-tech products this involves integrating localization throughout a company, after proper internalization and product design, as well as marketing, sales, and support in the world market.*
- *Internalization is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design. Internalization takes place at the level of program design and document development.*
- *Localization involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold.*
- *Translation is only one of the activities in localization; in addition to translation, a localization project includes many other tasks such as project management, software engineering, testing, and desktop publishing. (Lingo Systems & ATA, 2002: 4)*

As a conclusion, a common factor between all the definitions above stated is the fact that they all agree that localization is strongly connected with digital content translation.

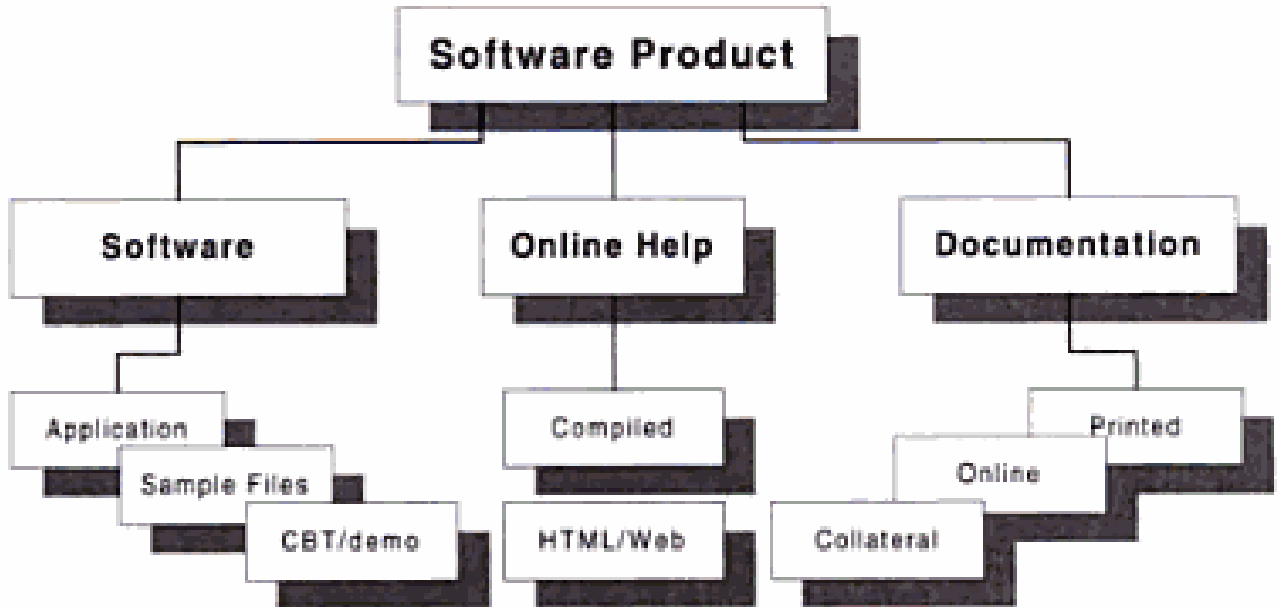


Figure 5. Esselink's example of localization products

Localization is needed for translating software, smart phone applications, web pages, multimedia products (DVDs, CDs, GPS, cameras, etc.), videogames, etc.

This industry is continuously flourishing, therefore there is a powerful competition on the market and companies need to localize their products in many languages because a local market is always stronger than a global one. Although most of the content that needs to be localized is released in English, big companies prefer to pay for this process because it means they can conquer external markets.

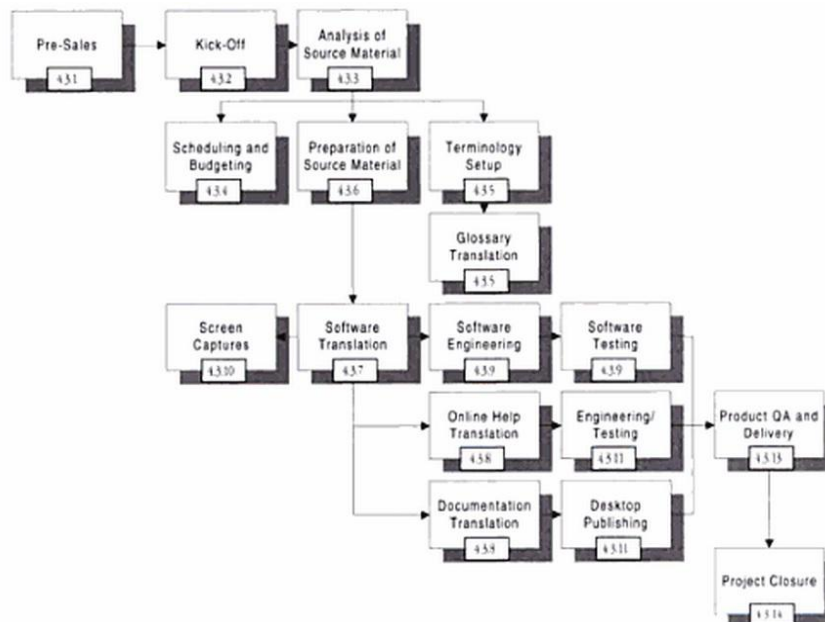


Figure 6. Esselink's example of the localization process

As we can observe in Figure 5, localization is a complex process that implies high costs; for this reason, it is not profitable for small companies to adopt this strategy and much of the content remains unlocalized. The financial aspect is one of the main reasons for which the products need internalization, because:

High costs and numerous cultural problems result from just taking home a product (usually software produced for the US market) and directly translating its natural-language strings into another language. That binary translation interface has to be repeated for each new language; it sends costs and spiraling. Overall expenses are lower if the home product is first prepared for all or some of its later localizations. (Pym, 2004: 31)

To give a specific example, we will take the case of the Spanish language and its variants in the Spanish speaking countries. To localize a videogame, taking into account all linguistic specificity for Spain, Mexico, Argentina, Peru, and Venezuela it could result very expensive, but when trying to use a homogeneous language which can be understood by a person living in Spain and someone from Peru or Venezuela, these costs are significantly reduced. "Thus organizations have to tailor their products to match the language and culture of the countries they intend to do business in, including countries with different varieties of the same language" (Quah, 2006: 20)

An important aspect to be mentioned is that the process of translation of a product to be localized often begins while the product is still elaborated. Also, "first one or more beta versions or release candidates of the product are translated before the final version is released to localization", therefore, the use of a localization tool is required, even though sometimes "additional edits or corrections are no longer stored in translation memory" because, usually, these changes are made before releasing the product and there is the risk of losing "all desktop publishing or testing work". In this case, the common practice is to update the translation memory separately. (Esselink, 2000: 22)

As regards to the localization CAT tools, these "are used in conjunction with other computer-aided translation tools such as translation memory systems and terminology-management systems", as stated above. (Quah, 2006: 20)

Kin Quah divides the localization CAT tool in management tools, which helps in keeping track of the complete localization project, and translation tools, the ones we have seen so far.

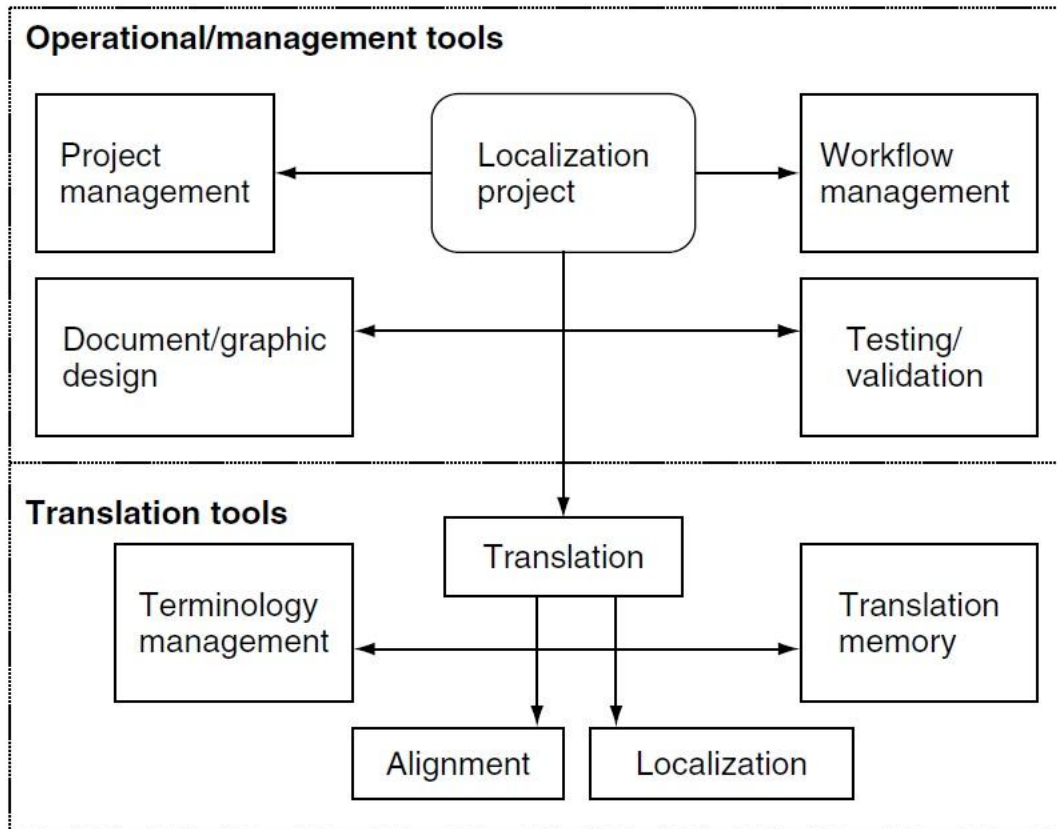


Figure 7. Kin Quah's example of tools used in localization

We have been focusing more on localization and its tools because the free tool we are about to submit to an exhaustive analyze, besides being very helpful for translating general and specialized content, can also be used to localize content.

As mentioned before, the terminology employed in localization is quite repetitive, therefore, this set of tools are essential in order to maintain the quality and the efficiency of the translation process and in a very tight deadline, fundamental requirements in this process.

To complete, we must highlight the conspicuous connection between translation and technology, hence the need for proper training and continuous improvement of technology knowledge for translators so that they can keep up with the workflow between the two of them.

5. Free and open-source software

To return to a topic previously acknowledged, we said that there are several types of software in terms of license, price, and restriction of use and/or modification. We consider it is necessary to explain the difference between free and open source software, as we will see in detail a tool which is both free and open source. However, we need to clarify first what free software means. Richard Stallman, the founder of Free Software Foundation, highlights the fact that the word *free* refers to *freedom*, not to price. (FSF, 2009)

According to the Free Software Foundation, free software is defined by four freedoms:

- *The freedom to run the program for any purpose.*
- *The freedom to study how the program works, and adapt it to your needs.*
- *The freedom to redistribute copies so you can help your neighbor.*
- *The freedom to improve the program, and release your improvements to the public, so that the whole community benefits.*

There are the [Free Software Foundation](#) that represents a non-profit organization whose mission is to promote computer user freedom and to defend the rights of all free software users, and the [Open Source Initiative](#), a non-profit corporation whose purpose is to educate and inform about the advantages and benefits of open source software.

The main difference between free and open source software resides in the fact that the code source might be or not publicly available. Therefore, open-source software could be free of charge, or it could cost money, but anyone with the required skills could look at and change the source code and, later on, make those changes available for everyone. However, there is also free and open-source software (FOSS) which is under no restrictive copyright and whose source code is openly shared for improvement.

The most common FOSS licenses such as the GNU General Public License allow the user to:

- *Install the software on as many machines as he/she wants.*
- *Allow any number of people to use the software at once.*
- *Copy the software and give it to anyone*
- *Modify the software, as long as certain features are kept intact (most commonly the licensing agreement).*

- *Freely (in the sense of «without restrictions») and for any purpose distribute or sell the software without paying royalties to the original developer. (McKay, 2006: 96)*

Unfortunately, users are so accustomed to using commercial software and, most of them, are not even aware of what FOSS represent and, quite often, if they use FOSS, they are not even aware of using it:

En realidad, muchos de ellos utilizan programas informáticos como Mozilla Firefox, el gestor de vídeo digital VLC o el paquete ofimático OpenOffice.org sin ser conscientes de que son programas libres o, como mucho, sabiendo que son gratuitos y sin hacerse más preguntas al respecto. (Díaz Fouces & García González, 2008)

The United Nations Educational, Scientific and Cultural Organization outlines the importance of FOSS in our society describing FOSS as a "practical instrument for development as its free and open aspirations make it a natural component of development efforts in the context of the Millennium Development Goals". FOSS helps in the development of effective infostructures because of the "non-discriminatory standards for information handling and access". Conjointly, the FOSS model "provides interesting tools and processes with which people can create, exchange, share and exploit software and knowledge efficiently and effectively". (UNESCO, 2014)

Therefore, there are many advantages that this type of software offers and the translation community takes great benefit out of it. In addition, they are able to translate the software itself and personalize it and, moreover, distribute it, so that other translators could use it. In the case of commercial software, it has many restraints as to the number of computers in which we are allowed to install the software, the price, supported format, etc.

So, of particular interest to translators is that anyone is free to create a language-specific version of any piece of free and open source software; you can translate it yourself and release your translation to the public, as opposed to asking a proprietary software company to do this for you. The opposite of FOSS is proprietary software, which means software whose source code is owned by a person or company, and which is almost always sold or given away under a more restrictive End-User License Agreement (EULA). (McKay, 2006: 97)

As we explained at the beginning of this paper, technology skills are a requirement for the modern translators, and having at their disposal FOSS, they can reduce costs, time and increase productivity. They need to be aware of the fact that even if it is a free tool, it does not mean it is less useful than a commercial one; this might be the main reason for which translators are reticent in using them.

Free CAT tools as an alternative to commercial software: OmegaT

Also, while FOSS can be upgraded by volunteers (because it is open software) we have access to upgrade it with no cost at all, commercial software offers upgrades at a very high price and, actually, without too many improvements. (McKay, 2006: 98)

Moreover, translators are not conditioned anymore on using a specific CAT tool for a client or to share their TMs or TMSs because FOSS is improving and trying to use a unique format, as we said earlier.

In what concerns translation students and freelance translators, we must insist on the fact that these resources are extremely useful and help them save money. There are many compared studies of the total cost that translators should spend on software and all of them show that the price difference between commercial software and FOSS is colossal. Corrine McKay, in an article for the magazine Panacea, demonstrates this by comparing the prices from Amazon.com and Translationzone.com:

Proprietary Software	Open Source Software
Microsoft Windows XP Professional: \$262.99	Ubuntu Linux: \$0.00
Microsoft Office XP Professional: \$225.00	OpenOffice.org \$0.00
Microsoft Outlook: \$89.99	Mozilla Mail: \$0.00
Quick Books Basic: \$189.99	Gnu Cash: \$0.00
Trados 7 Freelance: \$895.00	OmegaT: \$0.00
Total: \$1,662.97	Total: \$0.00
	OR as above, plus:
	CrossOver Office Professional: \$74.95
	Microsoft Office XP Professional: \$225.00
	Wordfast: \$220.00
	OR as above, plus:
	Heartsome XLIFF editor: \$88.00

Figure 8. Panacea's comparison of commercial and open-source software costs

Additionally, another reason for which we consider it is important to introduce free CAT tools to students is because they will become professionals and they need to acquire experience managing these tools to participate in different free projects to gain experience:

[...] as Vidal and Filho will argue in their papers on free software localization, students, as future translators, should become aware of the advantages of getting involved in free software localization projects, first as an excellent means to accumulate some real translation experience, as part of the localization process that involves a good number of steps and agents, and then as a form to contribute their skills to projects that benefit them in the first place. (Díaz Fouces & García González, 2008)

This is why we consider important that free CAT tools, along with commercial ones, should be included in the teaching syllabus. They can familiarize with both, but they can choose which one adopts for their particular use, without being conditioned by lack of paucity and price. We are in favor of these tools because they are very useful when dealing with technical texts, business texts, and, generally, repetitive terminology texts.

CAT tools might become a problem when the translator shifts part of his focus from the outcome of his work on obtaining the required tool and learning how to use it. Usually, the educational institutions which have incorporated CAT tools in their syllabus use commercial tools in their training programs for creating "quasi-professional translation environments". The academic institutions manage to acquire educational licensees for CAT tools like MemoQ, SDL Trados, Déjà Vu, and Wordfast for both the institution and the students. An inconvenient would be that commercial CAT tools vendors change their policies very often and the institutions cannot assure a continuity of their training programs. (Cánovas & Samson, 2001: 49)

On the other hand, free CAT tools are available for everyone, therefore, institutions can access them include them in their training programs for students without having to deal with economic issues and changing policies. Free software possesses similar characteristics as commercial software and offers the same basic features, except that it is free. Besides, most of the time extremely complex tools that offer a diverse range of features might prove difficult to work with and be rather unhelpful for translators as Cánovas & Samson state below:

Generally speaking, they are less sophisticated but they may be extremely helpful in training contexts. (Indeed, this distinction is important across all educational fields, not just in translation. Sophisticated and expensive commercial applications may be state-of-the-art but at the same time offer far more functionality than is requires in a training context.) (Cánovas & Samson, 2001: 49)

Bearing these considerations, we strongly believe that free and open-source software "should be the preferred, fully valid, professional option for pre-service translator training, both in classroom and distance learning courses." (Cánovas & Samson, 2001: 49)

III. ANALYSIS

1. Introduction to analysis

The following part focuses on one particular CAT tool, which is both free and open-source integrated translation environment: OmegaT. Consequently, the analytic part concerns about explaining the functioning of this tool and stating its highlights. After this thorough analysis, we will make a comparison of several aspects such as integrated tools, translation process, support file formats, system requirements, and some other characteristics.

The information presented below is gathered from the user's manuals of OmegaT, translation websites, and our conclusions after using the software to translate the webpage from FutureLearn *How it works*, distance learning and online education platform.

2. OmegaT - integrated translation environment

OmegaT is a free and open-source single-layer computer-aided translation tool, which can be redistributed and/or modified under the terms of the GNU General Public License. The software it does not come with warranty, as its purpose is to help the translation community and encouraging it and everyone that may find it useful to participate to its improvement.

As we said before, OmegaT is an open-source program; hence, we can modify the source code. In order to be able to build it from source, we need to download the Ant¹ program, which can be found here: <http://ant.apache.org/bindownload.cgi>.

2.1. Project properties

The way OmegaT manages a translation job is by organizing it in different folders for its most important features gathered in a project.

To set up a project in OmegaT is a very easy process, as we only need to follow the steps through the assistant.

¹ *Ant* is a programming language designed to build applications from the source code in Java.

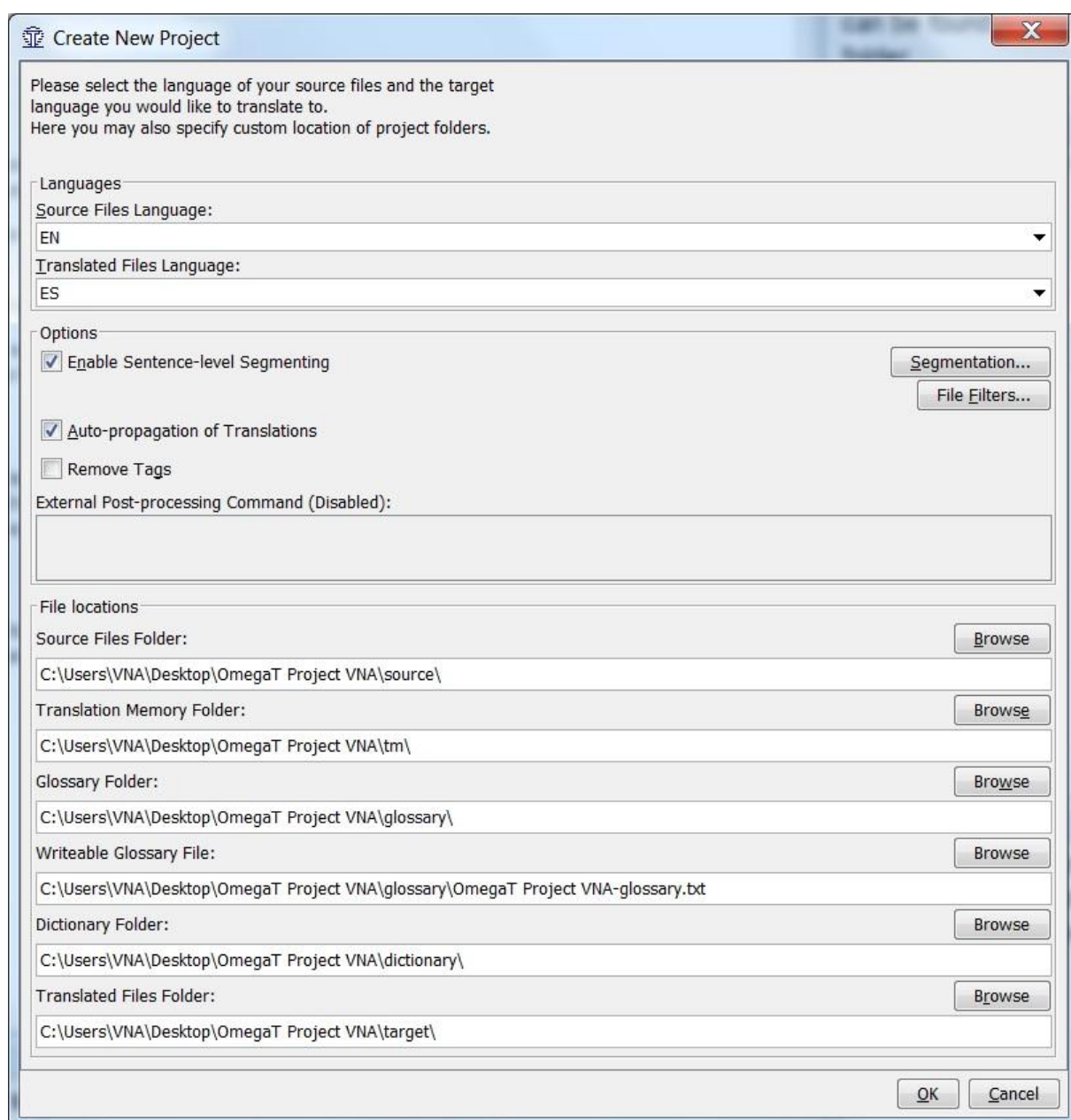


Figure 9. Project properties window

First we need to select the language combination we are about to use in our new project. We consider important to mention that, besides the most common language pairs, OmegaT offers a plethora of languages from Afar to Zulu.

Secondly, as to the segmentation settings, OmegaT comes by default with the *Sentence-level Segmentation* check box enabled. This means the source files are divided into sentences to facilitate the editing process. This option is usually the most common one, as almost all CAT tools have the same segmentation settings. Nevertheless, OmegaT gives us the possibility to choose segmentation by paragraph, but if we choose this option, we cannot switch to sentence-level segmentation because the internal TM of our project will not be changed.

Changing segmentation settings may cause some already translated segments to be split or merged. This will effectively return them to the "untranslated" status, as they will no longer match segments recorded in the project memory, even though their original translation is still there. (Smolej, 2013: 36)

Thirdly, the auto-propagation check box is used for automatic translation when, in the source text, there are non-unique segments.

If checked, the first translated segment will be assumed as the default translation and its target text will be automatically used for later hits during the translation process. Mistranslated segments can of course be corrected later manually using Create Alternative Translation. If the Auto-propagation check box is not checked, the segments with alternative translations are left untranslated until the user has decided which translation is to be used. (Smolej, 2013: 36)

Fourthly, as regards to the *Remove tags* check box, when enabled, OmegaT eliminates all the tags from the source text. This means that the text is left without format, very useful option when working with text whose format is not important, such as converted PDF files, for example.

Lastly, in the *File location* section, we can choose where we want OmegaT to save all the folders and subfolders for the current project we are working on. If we enter names of folders that do not exist yet, OmegaT creates them for us. In case we decide to modify project folders, this will not move existing files from old folders to the new location we have chosen.

As regards to the file format, OmegaT is compatible with many formats such as XLIFF, ResX files, WiX Localization, Flash XML export, Wordfast TXML, QuarkXPress CopyFlowGold, etc.

2.2. Editing field behavior

As many stand-alone CAT tools, OmegaT is working with segments; we can insert our translation into the correspondent source segment. After inserting our source text, OmegaT will automatically divide it into segments and leave us the option to insert the target text, to remove it afterwards if we want to or set a segment as empty by right clicking in that segment and selecting *Set empty translation*:



Figure 10. How to set an empty translation segment

Another very important feature of OmegaT is the fuzzy match. When translating, we have the option to insert the best fuzzy match, which is shown in the *Fuzzy Matches* pane. Basically, a fuzzy match is a similar segment to our current source segment.

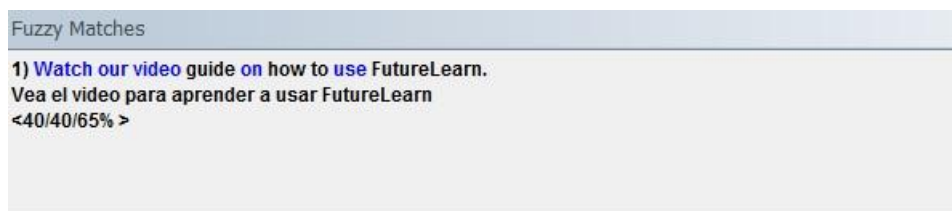


Figure 11. Example of how the text is displayed in the fuzzy matches pane

The selected fuzzy match is highlighted in bold; words that are missing in the segment we are translating are colored blue. If we are satisfied with the translation suggestion from the fuzzy match pane, we can select it with CTRL+1, in this case. The suggestions are numbered because it is easier to insert them with a simple hotkey combination, so we do not have to interrupt the translation process. In the moment when we decide to move to the next segment, usually the tool automatically selects the first fuzzy match, but we may choose a different one by pressing CTRL and the correspondent number.

In the example above, the source segment is *How it works* and the top match is 100% because all the words match.

2.3. Translation memory

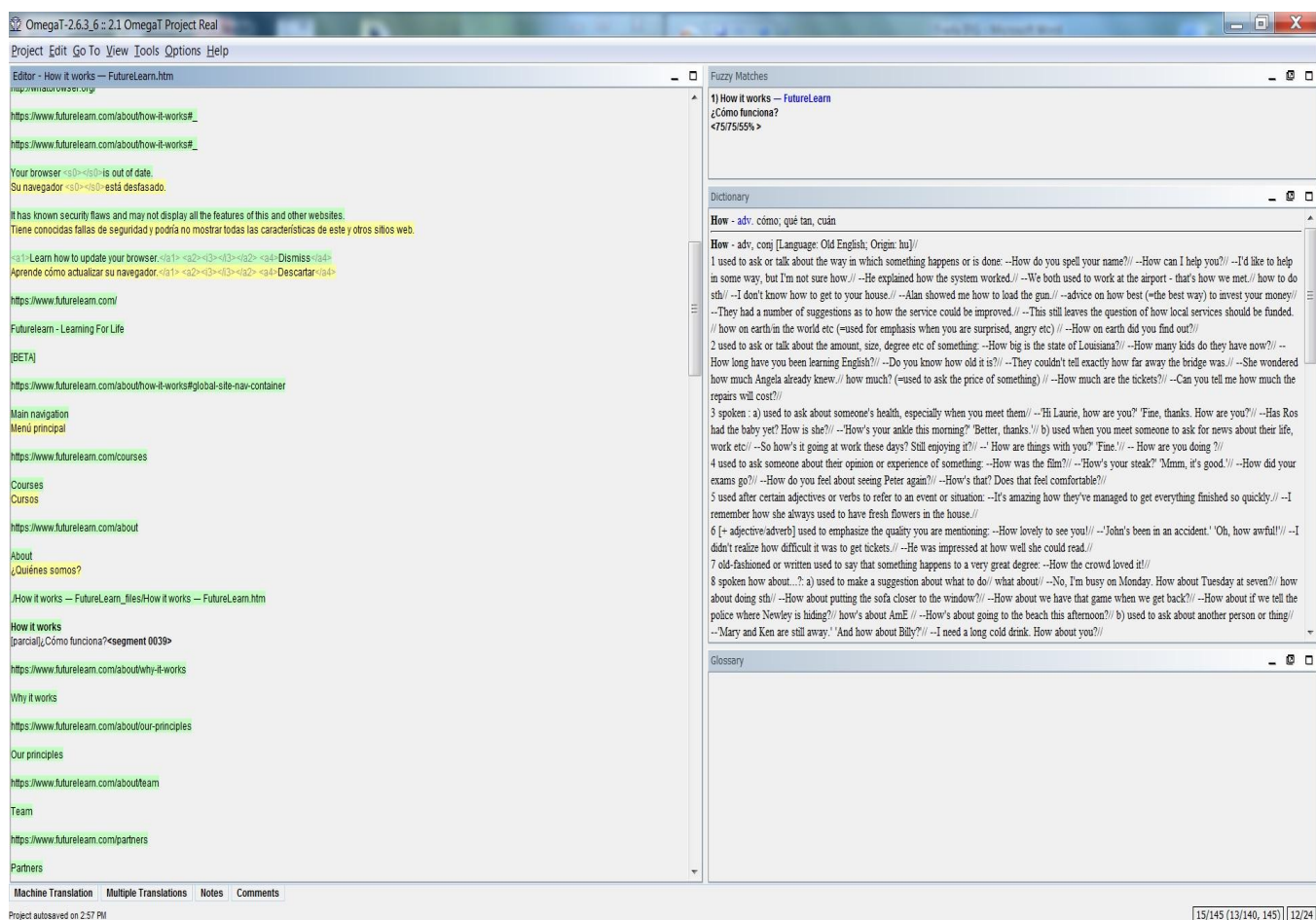


Figure 12. Fuzzy matches pane

The *Fuzzy Matches* pane can be easily customized in the menu *Option, External TMXs* as shown below:

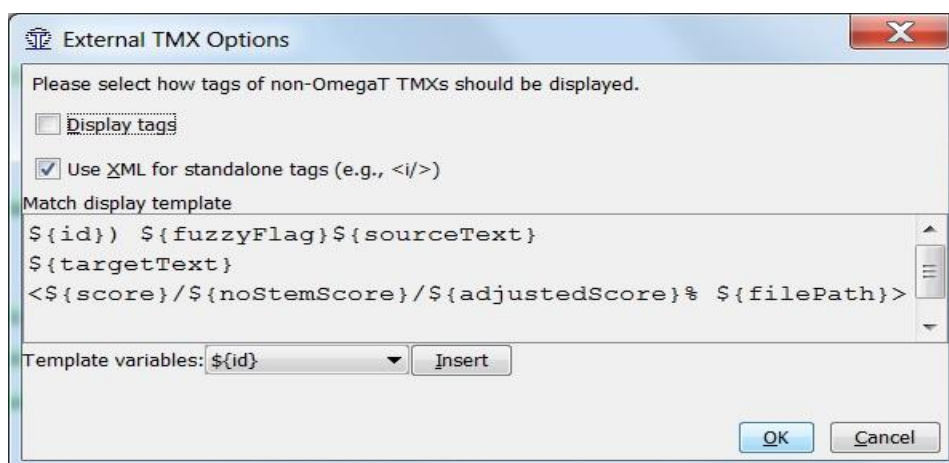


Figure 13. External TMX options

In the figure above, we can visualize the default settings for *External TMX Option* in OmegaT, but the context can be personalized using some variables that are available in the

OmegaT user's manual. Therefore, we do not need advanced programming skills, as we can find everything explained as below:

<code>#{id}</code>	Number of the match from 1 to 5
<code>#{sourceText}</code>	Source text of the match
<code>#{targetText}</code>	Target text of the match
<code>#{diff}</code>	String showing the differences between the source and the match. Hint: use this if the text that you are translating has been updated.
<code>#{score}</code>	Percentage with tokenizer
<code>#{noStemScore}</code>	Percentage without numbers and tags
<code>#{adjustedScore}</code>	Percentage adjusted
<code>#{fileNameOnly}</code>	Name of the TMX
<code>#{filePath}</code>	Full path of the TMX
<code>#{fileShortPath}</code>	Path of the TMX starting from the root of /tm
<code>#{creationID}</code>	Author of the match
<code>#{creationDate}</code>	Date of the match
<code>#{fuzzyFlag}</code>	Indicate that this match is fuzzy (currently only for translations from PO files with the #fuzzy mark)

Figure 14. Match pane setup

The OmegaT user has the option to select what text we desire to be inserted into a segment that is not translated yet. We can either leave the segment empty or introduce the source text.

As regards to the fuzzy match, we can choose to insert the best match and set the minimal similarity. In addition, when a fuzzy match is inserted, at the beginning of the segment, OmegaT warns us that the translation inserted is partial by intentionally leaving the word *[partial]* until we choose to delete it after editing the segment, as we can see in Figure 15. We find this

feature particularly interesting and useful because it does not allow us to forget to post-edit the segment.



Figure 15. Fuzzy match pane

We can choose other settings in *Editing Behavior Options*:

- Attempt to convert numbers when inserting a fuzzy match
- Allow translation to be equal to source
- Export the segment to text files
- *Go To Next Untranslated Segment* stops when there is at least one alternative translation

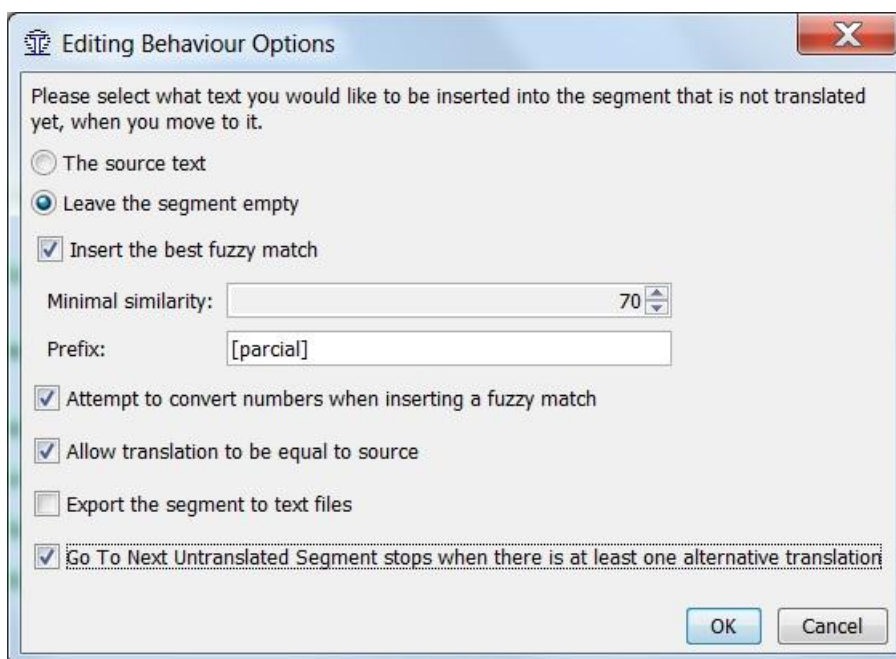


Figure 16. Editing Behavior Options dialogue window

Translation memories in OmegaT could be located in *omegat* folder, main project folder, *tm* folder, *tm/auto* folder, *tm/penalty-xxx* folders or wherever we may desire to have additional TMX. While we translate, OmegaT saves all our work automatically in *project_save.tmx* and all of subfolders, so we do not have to worry about losing our work.

We may also notice that if the target segment is identical to the source segment, OmegaT chooses not to register our translation, but this does not mean we have to post-edit the target text. Those segments are also automatically transferred to the target text.

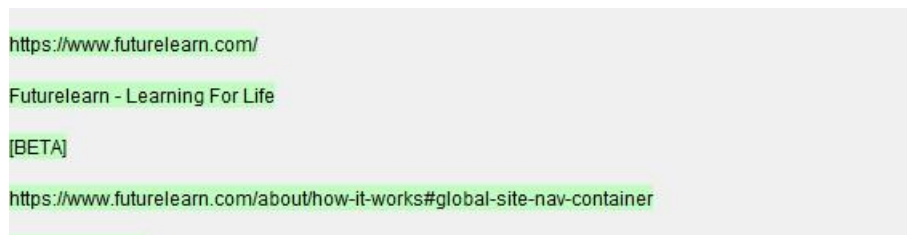


Figure 17. Segments automatically transferred to the target text

All the segments translated in a project are saved as well, even if we delete a source segment, some matches may still appear because they were recorded before having done any changes. Those segments saved in the TMX are called orphan segments or orphan strings and they could appear in the *Match Viewer*.

2.3.1. Manipulating TMs

As we said before, translation memories are very useful tools because they gather all the translated segments with their matching source ones. OmegaT allows us to recycle an existing TM created with the program itself or with another CAT tool.

Therefore, if we have a new translation job that might be similar in terms of terminology or phraseology with an older one, we can use one of the OmegaT TMs. There are few easy steps we need to follow in order to be able to take advantage of our previous work. Firstly, we need to copy one of the three TMX files (*project_save.tmx*) automatically created by OmegaT, which could be located in the project folder *omegat* (unless we have chosen another route to save our TM). Secondly, once located, we need to copy that file and paste it in our TM folder of our new project. Now OmegaT will suggest us possible translation units in the fuzzy matches viewer.

Some OmegaT users choose to create a TMs storage folder where they gather all TMs for a certain client or on a certain topic and then link that folder to the new project they are working at. We think it is an excellent opportunity that OmegaT offers to its users because it is a real time saving, as we do not need to merge TMs we can place as many as we like in the correct folder. OmegaT's user manual advises that there is a small inconvenience though: if we have too many TMs, it might slow down a bit OmegaT.

If we have worked with previous versions of OmegaT, we have the possibility to upgrade older TMX files in order to increase fuzzy matching quality.

In regards to importing TMs created with different CAT tools than OmegaT, we should make clear that, although it supports both level 1 and level 2 TMX files, OmegaT is not fully compatible with level 2 TMX, hence there will not be very high quality fuzzy match results.

For working in collaboration with other translators, we also have the option of sharing our TM. If we do not want to share the complete project, we can create a new one and copy the translation memory in the *tm* folder.

OmegaT also recognizes sources with existing translations if they are located in *project_save.tmx*. There is also the possibility to pre-process segments before translate them. For example, if we wish to create a pseudotranslation for testing purposes, OmegaT allows us to create an additional TMX file that contains all segments of the project. The translation in this new created TMX can be:

- Translation equals source (default).
- Translation segment is empty.

The TMX file can be given any name we choose to designate. A pseudo-translated memory can be generated if we insert the following command line parameters: `java -jar omegat.jar --pseudotranslatetmx=<filename> [pseudotranslatetype=[equal|empty]]` (Smolej, 2013: 64)

2.4. Terminology management

2.4.1. Dictionaries

OmegaT allows us to download and install dictionaries such as *Merriam Webster*, *Longman Dictionary of Contemporary English*, *The Britannica Concise Encyclopedia*, etc. On the [StarDict Platform](#) we can look up for our language pair combination and download the dictionary. After downloading we have to decompress the file and make sure that it is located in the folder *dictionary* (see Annex 2).

2.4.2. Glossaries

We can also add glossaries to help us through the translation process. We have the possibility to create one and add entries while translating a file or we can use existing ones.

When creating a project, a glossary folder is created automatically. If we want to add entries to the glossary, we can either right click on a term in the text editor and select *Add glossary entry*, use the shortcut CTRL+SHIFT+G, or click *Edit, Add glossary entry*. It is important to remember to save the changes, so that OmegaT recognizes the new entries.

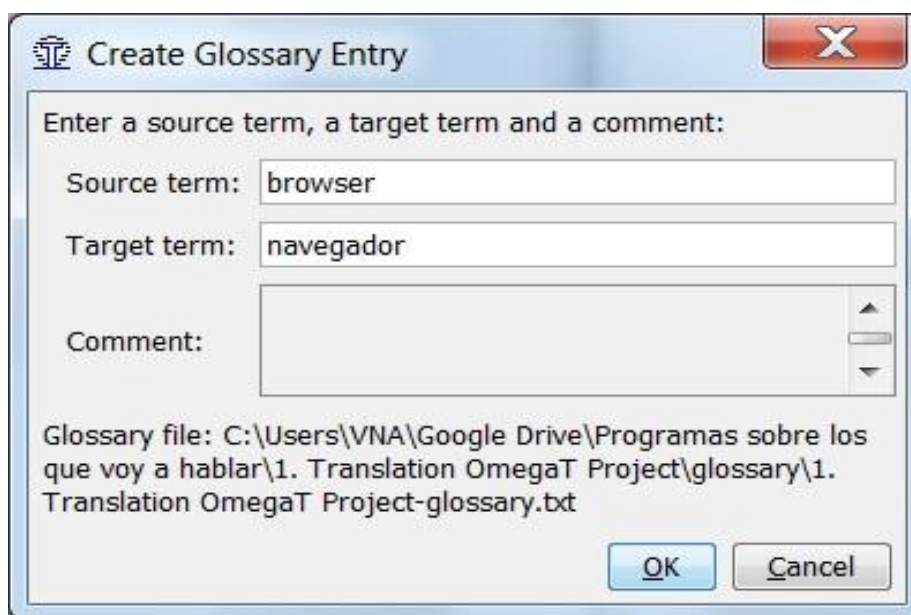


Figure 18. How to create a glossary entry dialogue window

In regards to the glossary format, usually it supports simple plain text files with the extensions .tab or .utf8, but TBX (Term Base eXchange) format is also accepted, so we can exchange terminological data.

In addition, OmegaT is compatible with the commercial TMS Trados MultiTerm. We can import the file that MultiTerm generates and could be either .tab or TBX, in the case of recent versions of MultiTerm.

Another interesting aspect to mention is that, while translating, if there is a term in the segment that we are editing which is in the dictionary or in the glossary, OmegaT displays the translation suggestions in the Glossary pane, but only if the term is identical, because no inflected forms are recognized:

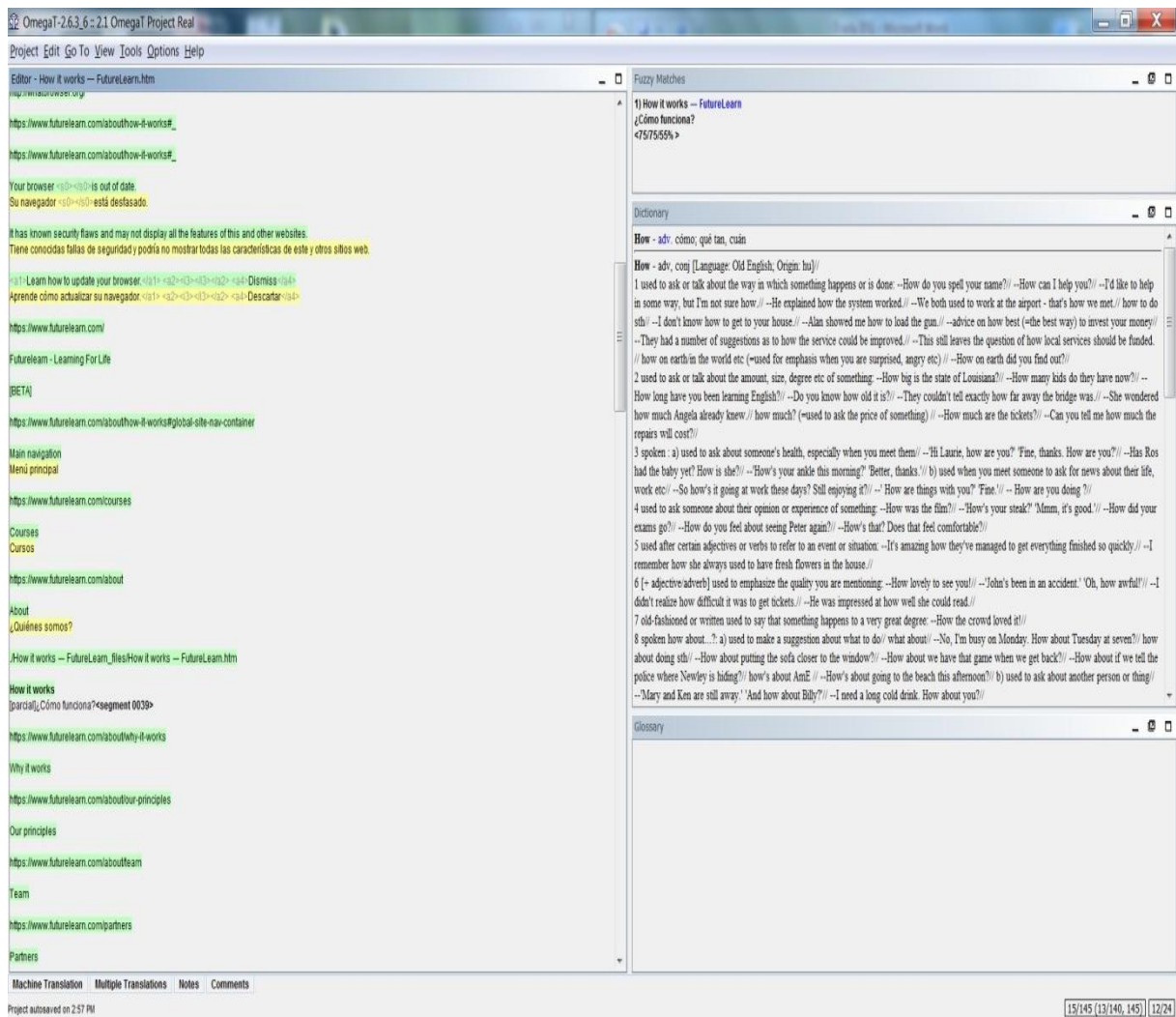


Figure 19. Glossary pane suggestions dialogue window

2.4.3. Spell checker

In order to use an OmegaT integrated spell checker, we need to install a spelling dictionary based on the spelling checker, which is used in *Apache OpenOffice*, *Thunderbird*, *Firefox*, and *LibreOffice*. We can either add a dictionary to a folder previously created, or we can install one because OmegaT offers the option of installing a new one by selecting *Install new dictionary*.

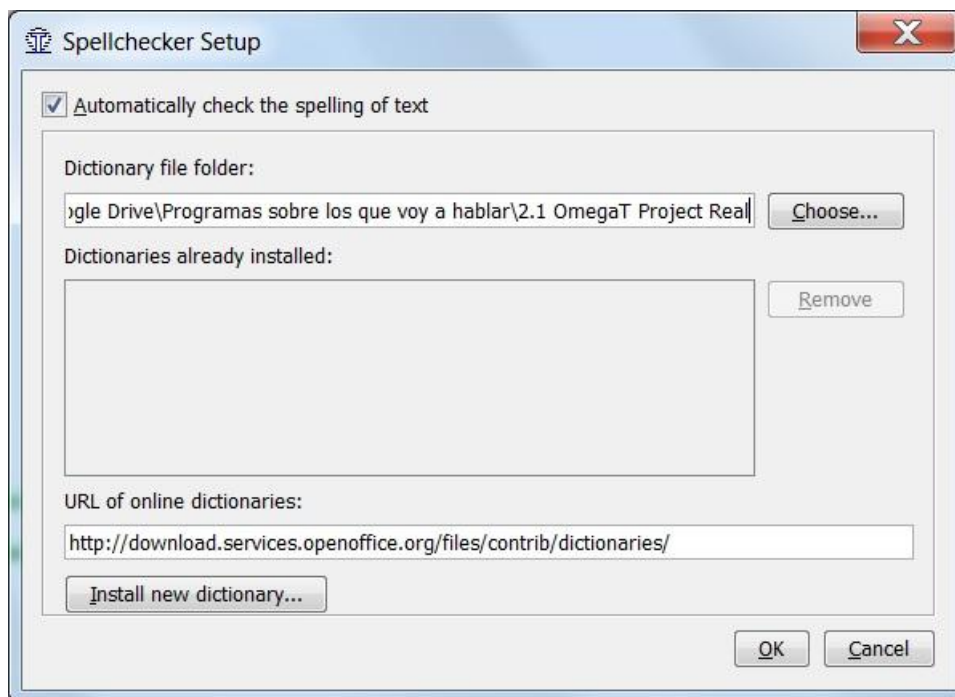


Figure 20. Installing a dictionary dialogue window

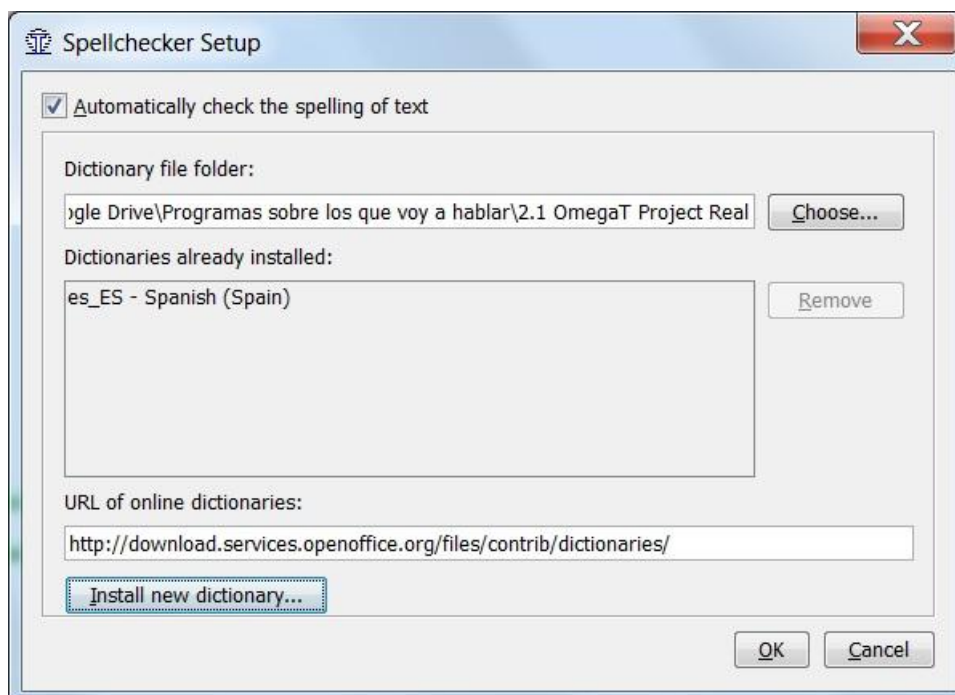


Figure 21. Display of the already installed dictionaries dialogue window

By enabling the *Automatically check the spelling of the text* box, OmegaT will check spelling while we translate. This is very efficient and time-saving because we do not have to come back to that segment to correct errors.

2.5. Machine translation

OmegaT offers us the possibility of using a machine translation (MT) such as *Google Translate*, *Belazar*, *Apertium*, and *Microsoft Translator*. The quality of the translation suggested by these machines translation has nothing to do with OmegaT, as they are independent tools.

Moreover, not all of them are free tools. *Google Translate v2*, for example, is not a free tool and, before using it, the application requests billing information.

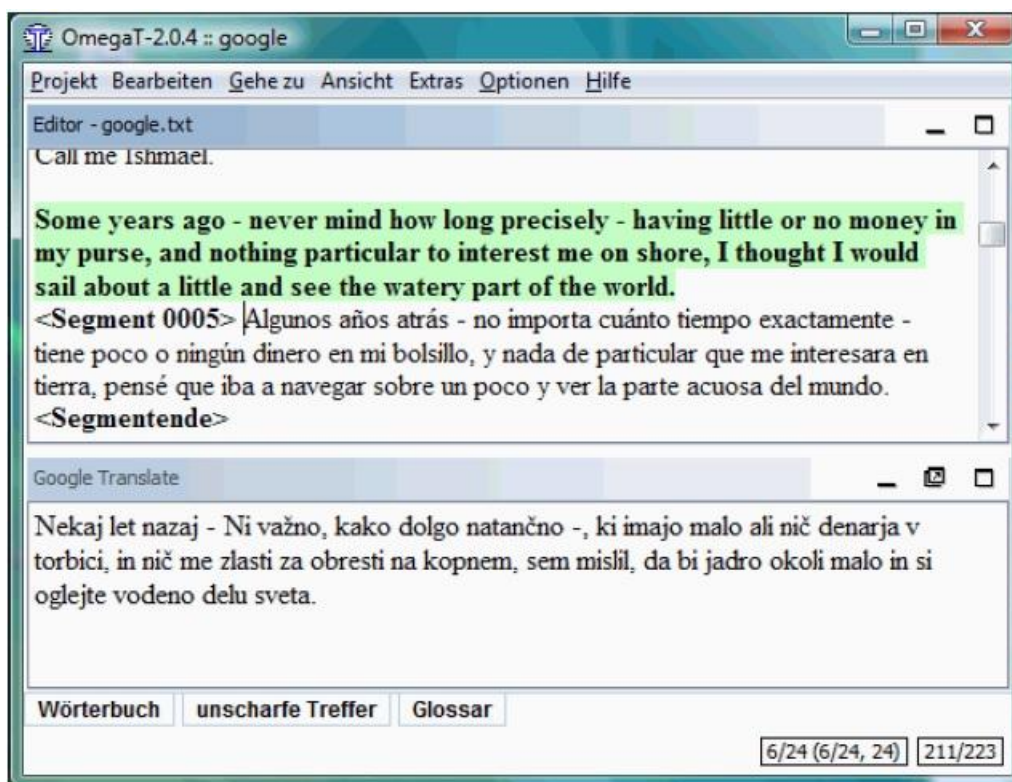


Figure 22. Google Translate behavior window

Google Translate offers a great variety of language combinations: 50 languages. *Belazar* offers Russian-Belarusian language combination, *Apertium* is increasingly implementing more language combination, but so far it offers CA, ES, GA, PT, OC, FT, and EN.

We are not obliged to use a MT, but if we use it and we are not satisfied with the outcome, we are always free to edit it. We are not able to use any of the MT mentioned above if we are not connected to the internet.

2.6. Automatic Java Properties Aligner

In console mode, that is, using the tool in a scripting environment such as Java, OmegaT is able to align documents as long as we possess both the source and the target text. After the alignment, OmegaT will create a TMX file that can be found in the folder named *align.tmx*.

Additionally, it allows incorporation of external alignment tools such as *WinAlign*.

2.7. Plugins

Another great advantage of OmegaT is the possibility of installing many useful plugins to extend the basic functionality of the tool. Here are some examples:

2.7.1. Tokenizer plugin

The tokenizer or stemmers is a plugin which provides an improvement of the fuzzy and glossary matches. This tokenizer works on an algorithm that has the ability of recognizing variables of the word's root. Hence, even if the word is not identical, the plugin is able to find matches.

On Sourceforge webpage we can download it for free and install it. It is compatible with Mac OS, Windows and Linux. The installation process is the same as for all the other OmegaT plugins. We need to create a new folder called *plugins* in OmegaT installation folder and decompress the download the archive here. In addition, to make it functional, it is necessary to specify which tokenizer we wish to use for the source language and which one for the target language.

2.7.2. LanguageTool plugin

LanguageTool in OmegaT is a style and grammar checker software available in many languages. We can download it from the [LanguageTool](#) webpage and install it on Windows or on Linux, as it is compatible with both systems. In order to make the plugin functional, we need to create a new folder called *plugins* in OmegaT installation folder and decompress the download the archive here. Now we only need to restart OmegaT to make the changes effective and enable the option *Language Checker*.

This plugin is packed with many language rules and, when one is triggered, that sentence in the editor will be underlined in blue and, if we move the mouse over that sentence, it will appear the explanation.

It is interesting to mention that this checker is also open-source software, which means anyone can add rules for the existing languages and improve it as we can see in Figure 23.

Free CAT tools as an alternative to commercial software: OmegaT

Language	XML rules	Java rules	False friends	Rule Maintainers
Asturian	61	0	0	Xesús González Rato
Belarusian	7	1	0	Alex Buloichik - Looking for co-maintainer
Breton	653	2	0	Dominique Pellé , Fulup Jakez
Catalan Variants for: Catalan, Valencian	1914	11	4	Ricard Roca, Jaume Ortola
Chinese	766	0	0	Tao Lin - Looking for co-maintainer
Danish	78	0	6	Esben Aaberg, Henrik Bendt - Looking for co-maintainer
Dutch	337	4	19	OpenTaal
English Variants for: Australian, Canadian, GB, New Zealand, South African, US	1077	15	326	Marcin Miłkowski , Daniel Naber - Looking for new maintainer
Esperanto	343	0	0	Dominique Pellé
French	2450	1	6	Dominique Pellé , Agnes Souque, Hugo Voisard (2006-2007)
Galician	153	2	113	Susana Sotelo Docio - Looking for co-maintainer
German Variants for: Austria, Germany, Swiss	1879	15	95	Jan Schreiber, Markus Brenneis, Daniel Naber
Greek	20	2	0	Panagiotis Minos
Icelandic	39	0	0	Anton Karl Ingason - Looking for new maintainer
Italian	134	2	37	Paolo Bianchini
Japanese	43	0	0	Takahiro Shinkai - Looking for new maintainer
Khmer	31	5	0	Nathan Wells
Lithuanian	4	1	0	Mantas Kriaučiūnas - Looking for new maintainer
Malayalam	18	1	0	Jithesh.V.S - Looking for new maintainer
Polish	1210	5	139	Marcin Miłkowski
Portuguese Variants for: Brazil, Portugal	146	3	0	Marco A.G. Pinto - Looking for a maintainer for Brazilian Portuguese
Romanian	458	4	0	Ionuț Păduraru - Looking for new maintainer
Russian	317	5	21	Yakov Reztsov
Slovak	55	2	0	Zdenko Podobný
Slovenian	85	1	0	Martin Srebotnjak
Spanish	97	0	28	Juan Martorell
Swedish	26	1	5	Niklas Johansson - Looking for new maintainer
Tagalog	44	0	0	Nathaniel Oco , Allan Borra
Ukrainian	191	5	0	Andriy Rysin, Maksym Davydov

The number of Java rules listed is only the number of rules specific to that language. There are some rules that deal with punctuation and that apply to almost all languages.

Figure 23. Rules in LanguageTool 2.6

2.7.3. Scripting plugin

The scripting plugin allows us to run scripts that are written in different scripting languages in OmegaT.

On Sourceforge webpage we can download it for free and install it. It is compatible with Mac OS, Windows and Linux. The installation process is the same as for all the other OmegaT plugins. After we installed it, a new submenu will be created in *Tools - Scripting*. In the *Scripting* dialog we can choose to load one for our current project. These scripts work as macros and, once we load them, they scan through all the segments of our current project to find if there is a translation for them and then it will print out both the source and the target of those segments.

As we said before, we are free to change the OmegaT source code. We can help OmegaT not only by translating the interface or the manuals, but we can also get involved in OmegaT team

projects. There are some steps we need to follow as project manager in order to create the project and give access to other members. We also need to be familiarized with *Apache Subversion* and *Git*, version control systems supported by OmegaT.

The advantages of a VC [Version Control] system for a team of translators can be summarized as follows:

- *Several team members can work on the translation project simultaneously without interfering with each other.*
- *They can share common material, like project translation memory and its glossary.*
- *Every three minutes by default, an updated version of data shared is available to the rest of the team.*
- *The system maintains versioning for data shared.*
- *Conflicts - for instance alternative translations of the same segment or glossary entry - can be monitored, resolved and merged. (Smolej, 2013: 96)*

Additionally, there is another way of improving our fuzzy and glossary match results. We can download and install a stopword list, which removes prepositions, conjunctions and other less important elements from the token stream.

3. Impressions

After examining the tool, we can say that OmegaT represents an excellent free and open-source CAT tool, which is worth introducing in our workstation.

Starting with the user interface, which is composed in two main sections - the editor in the left and the fuzzy match, the glossary, and the dictionary on the right, we can say that it is accessible, efficient, and enjoyable. It is very basic indeed, and if we have previously worked with a commercial CAT tool we might tend to underestimate the quality of the tool. However, once we get used to it, we actually find it very efficient because it has incorporated the main interactive aspects of the tool.

Another great strength of OmegaT is the way it generates all the source, translation memory, glossary, dictionary, and target subfolders. This way of organizing the project represents a good way of helping us being more organized as well. We can easily locate the folders if we wish to install glossaries, dictionaries, plugins, etc., retrieve the target text or manipulate TMs.

As regards to the file formats supported, we consider imperative to point out how well OmegaT deals with formatted text. We are not going to insist on plain text files supported, as we have already established that OmegaT is a very advanced CAT tool and, for being a free tool, it has improved considerably to the point that now it supports plenty formatted files.

We have tested the program by translating a HTML file and we were very satisfied with the outcome. It is a very good CAT tool for website localization, because we do not have to worry about tags or HTML language code as OmegaT maintains the structure and it displays only the translatable text, even though it saves all the non translatable files. Later on, we will reveal the results of our translation exercise.

We are aware of the fact that the glossary function is a bit poor, but we have to keep in mind that OmegaT is not designed to meet specific needs. In addition, its appearance could be improved, but, as far as we are concerned, the simple design leaves no place for distractions, therefore, it allows us to fully concentrate on the translation process. This could be an inconvenient for those who wish to personalize the font, the colors, etc., but the simplicity of operation is appreciated.

OmegaT is fast and responsive, easy to use, and most important it is free. It has many advantages as we will demonstrate in the following chapters. In fact, we can affirm that it is a very complete free CAT tool and it is constantly developing becoming a professional tool that can compete with commercial ones. The proof of this statement is last year's research project supported in part by the Spanish Ministry of Education and Science and by the Basque Government in which an improved version of OmegaT and its user guide were part of the publicly

Free CAT tools as an alternative to commercial software: OmegaT

available resources created within the project. (Gurevych & Kim, 2013: 4) Additionally, Marta García González has conducted a survey on the use of Free and Open Source Software and, regarding the use of CAT tools, "7 respondents out of 14 referred to OmegaT as their preferred CAT tool." (Díaz Fouces & García González, 2008)

Price	Free
Compatibility	<ul style="list-style-type: none">• Windows, Macintosh OS X, and Linux.• TMX format, which allows editing, importing and exporting• Plain text and formatted text
Translation Memory	<ul style="list-style-type: none">• Integrates TMs from other CAT tools• Supports various TMs simultaneously• Fuzzy Match feature• Pre-process segments before translating
Quality assurance	<ul style="list-style-type: none">• Allows spellchecker installation
Machine Translation	<ul style="list-style-type: none">• Google Translate• Belazar• Apertium• Microsoft Translator
Terminology management	<ul style="list-style-type: none">• Allows glossaries and dictionaries installation• Allows adding glossary entries while editing• Supports MultiTerm glossaries• Stopword lists
Alignment	<ul style="list-style-type: none">• Allows incorporation of external alignment tool

Miscellaneous	<ul style="list-style-type: none">• Allows plugins such as: Tokenizer LanguageTool Scripting• Allows contribution to OmegaT project• Allows modifying the source code
Technical support	<ul style="list-style-type: none">• Provided on a volunteer basis by OmegaT users

Figure 24. Compilation of OmegaT highlights

4. Translating with OmegaT

To examine the functions of OmegaT above recorded, we have translated FutureLearn's *How to* webpage. In this section we will concisely enumerate the steps we have followed for this translation process as the process of translating with this tool is pretty straightforward.

To start translating with OmegaT it was necessary to create a new project. To do this, from the *Menu* bar we select *Project*, then *New* and the following *Create a New Project* window appeared:

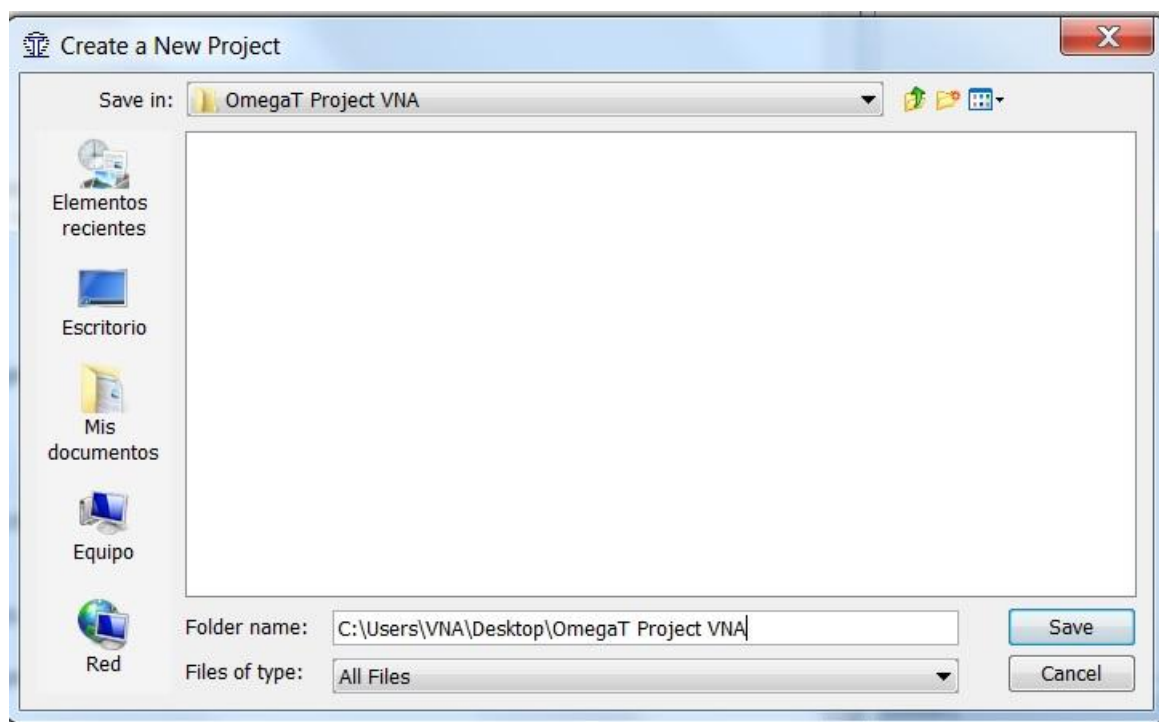


Figure 25. Saving a project window

We chose a folder we have previously created and then click on *Save* button. Automatically a new window comes into sight:

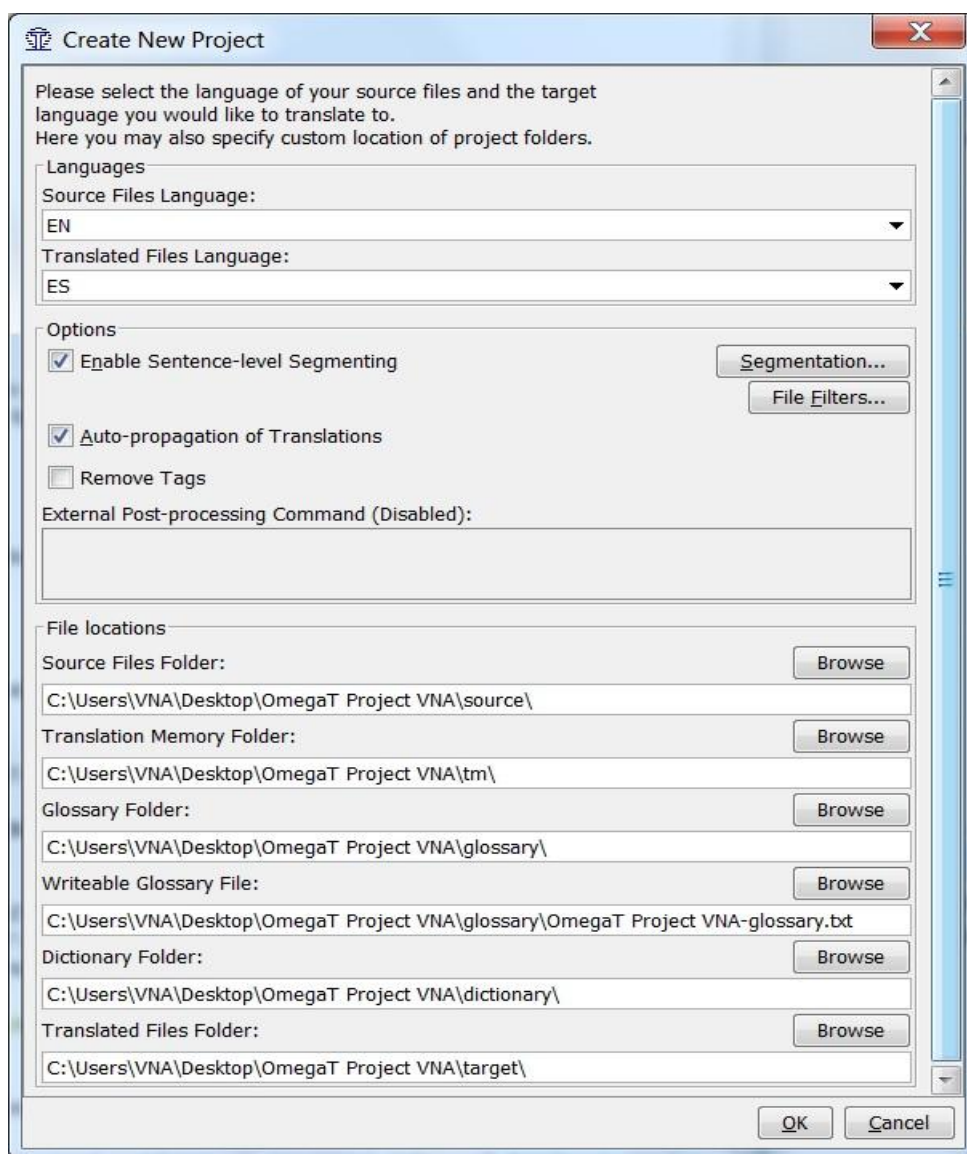


Figure 26. Create new project window

In this step we have chosen the language pair we needed for our translation, that is, English and Spanish. As regards to *Options*, we have used the default settings, as it was suitable for our project.

As we said before, OmegaT creates a main folder which contains five different subfolders: *source files*, *translation memory*, *glossary*, *dictionary*, and *target file* folder. The *Writeable Glossary File* refers to the file OmegaT generates when creating a glossary while translating.

To insert the file we have chosen to translate, we had to save the webpage in our computer and then import it in OmegaT.

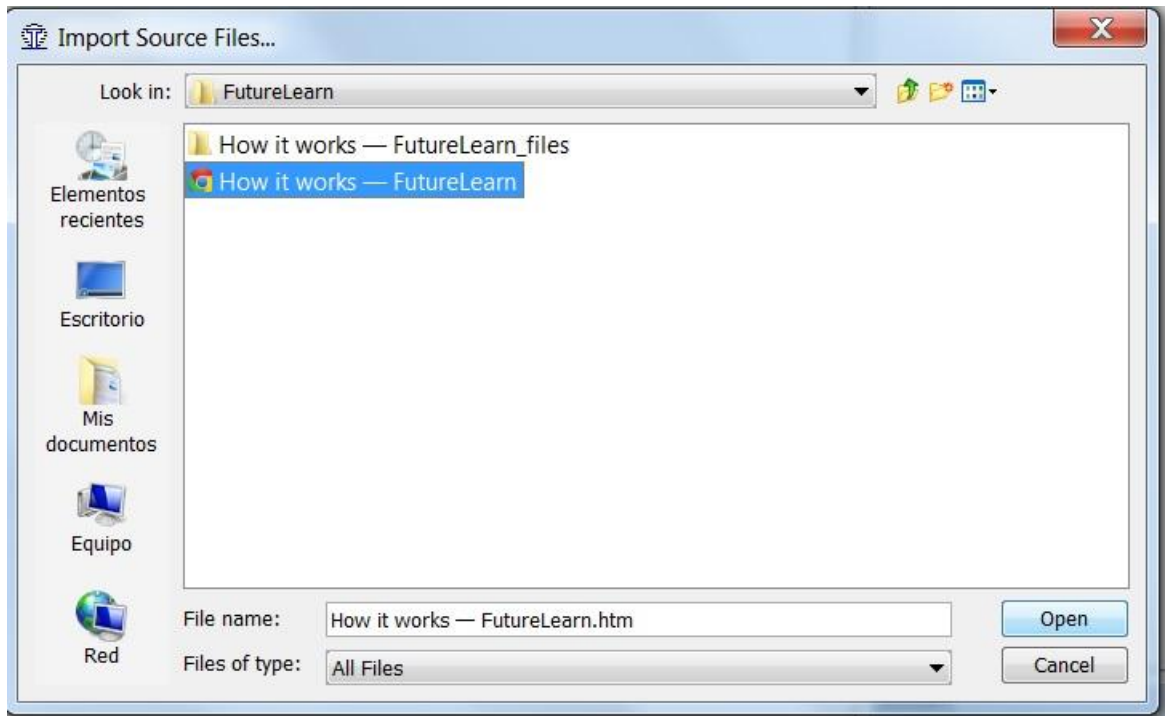


Figure 27. Import source files window

Before starting to translate, we have quickly personalized OmegaT in order to facilitate its manipulation:

- We added two dictionaries and a stopword list to the dictionary folder.
- We have configured OmegaT to mark the untranslated segments.
- We have selected the option to use *Tab* key to move to the next untranslated segments.
- We have changed the font.

In the following window we have inserted our translation segment by segment:

Free CAT tools as an alternative to commercial software: OmegaT

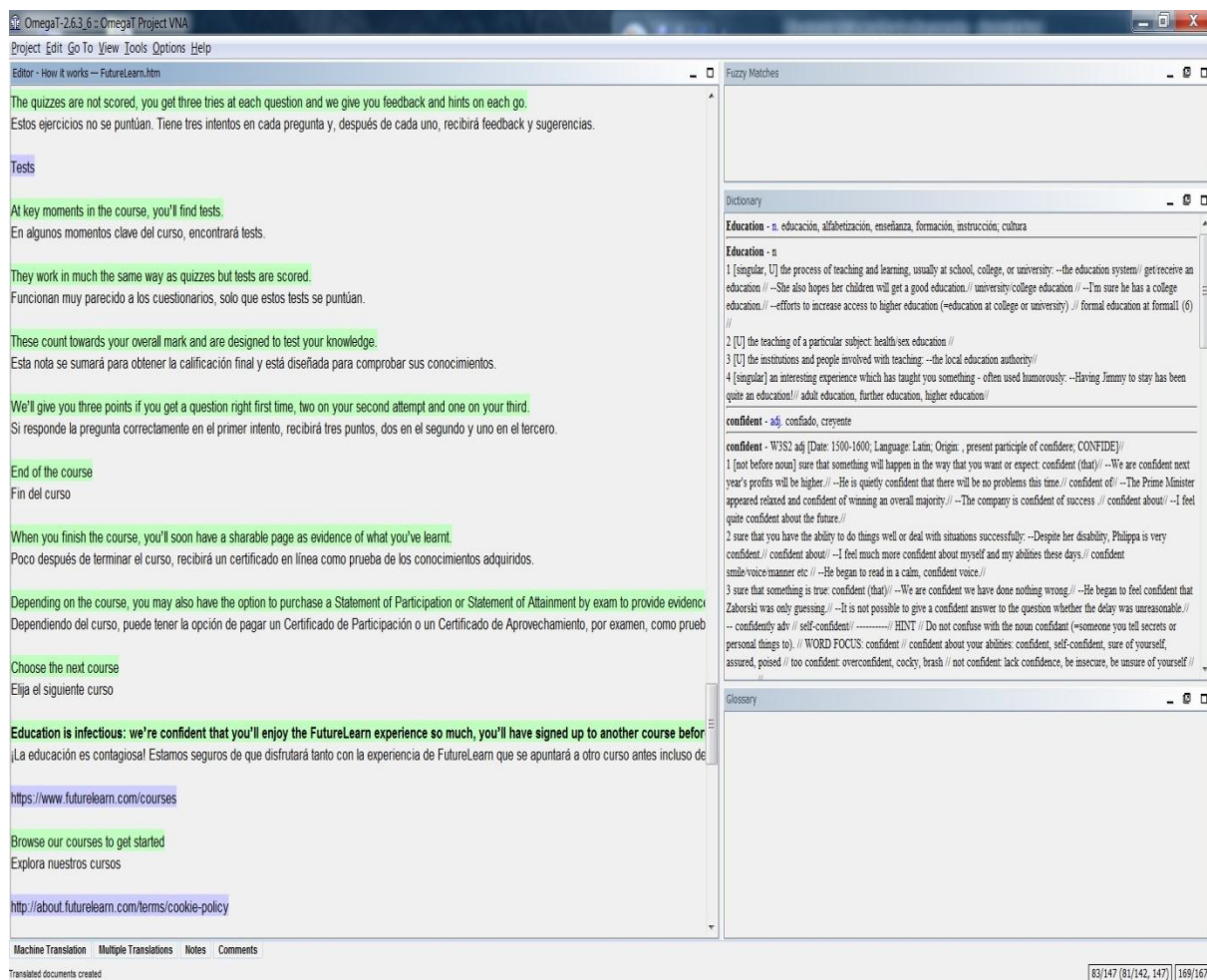


Figure 28. Editor window

While we were translating, the dictionaries we have installed offered suggestion to speed up the process, and, as we can see in the editor, both the source text and the untranslated text were marked with different colors.

When we finished translating, we have saved the project and selected *Create translated documents* from the *Project* menu.

To see the result of our translation, we have retrieved the HTML file from the folder target of our project as we can see below.

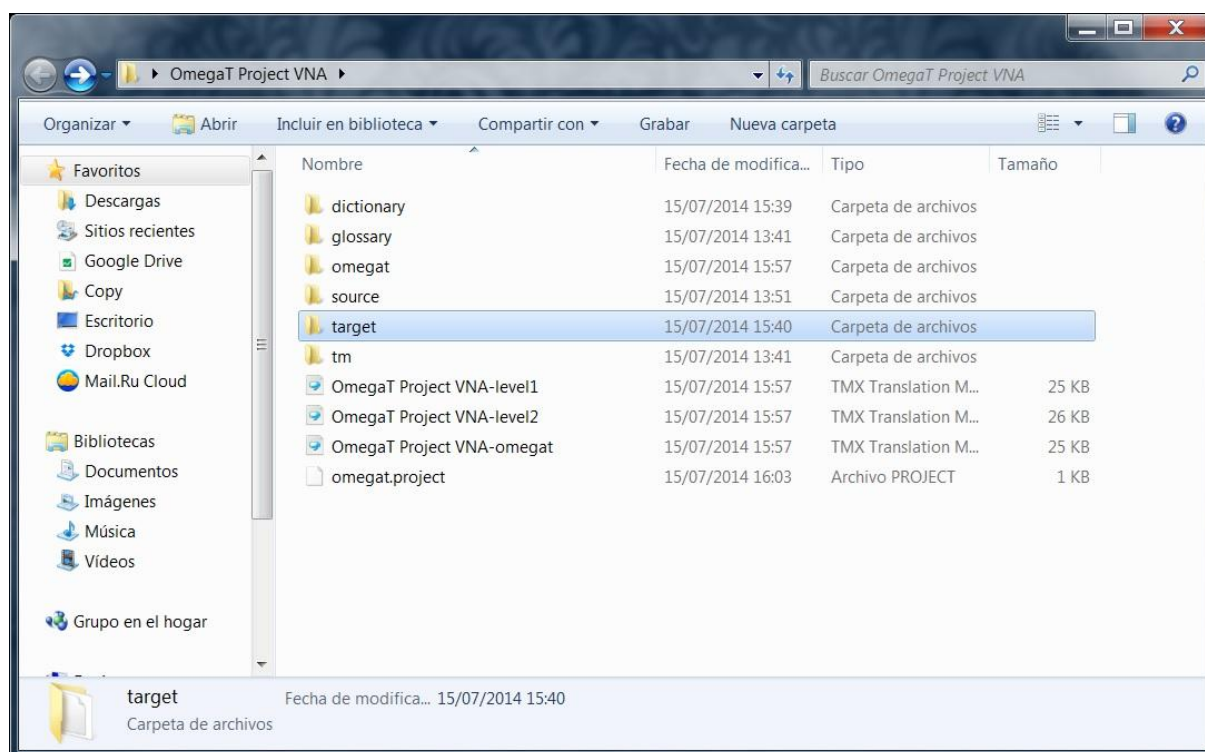


Figure 29. FutureLearn project folders

In the following section, as well as in Annex 3, we can see the final result of our translation exercise.

4.1. OmegaT results

We have chosen to translate with OmegaT FutureLearn's *How to* webpage to take advantage of OmegaT's ability to manage HTML files. Additionally, we have chosen this webpage because FutureLearn is a provider of free online courses and we support the idea of universal access to education and autonomous learning.

In regards to our translation, as we can see in the screenshots provided in Annex 3, the results are very satisfactory. The translated webpage is almost identical with the original, as OmegaT transferred all the tags successfully.

We have discovered a small inconvenience while we were translating: OmegaT does not inform us when we misspell a word. It is true we can run the Spellchecker when we finish translating, but we consider it would have been useful and time saving to highlight the word so we can correct it when we are working on that particular segment.

Free CAT tools as an alternative to commercial software: OmegaT

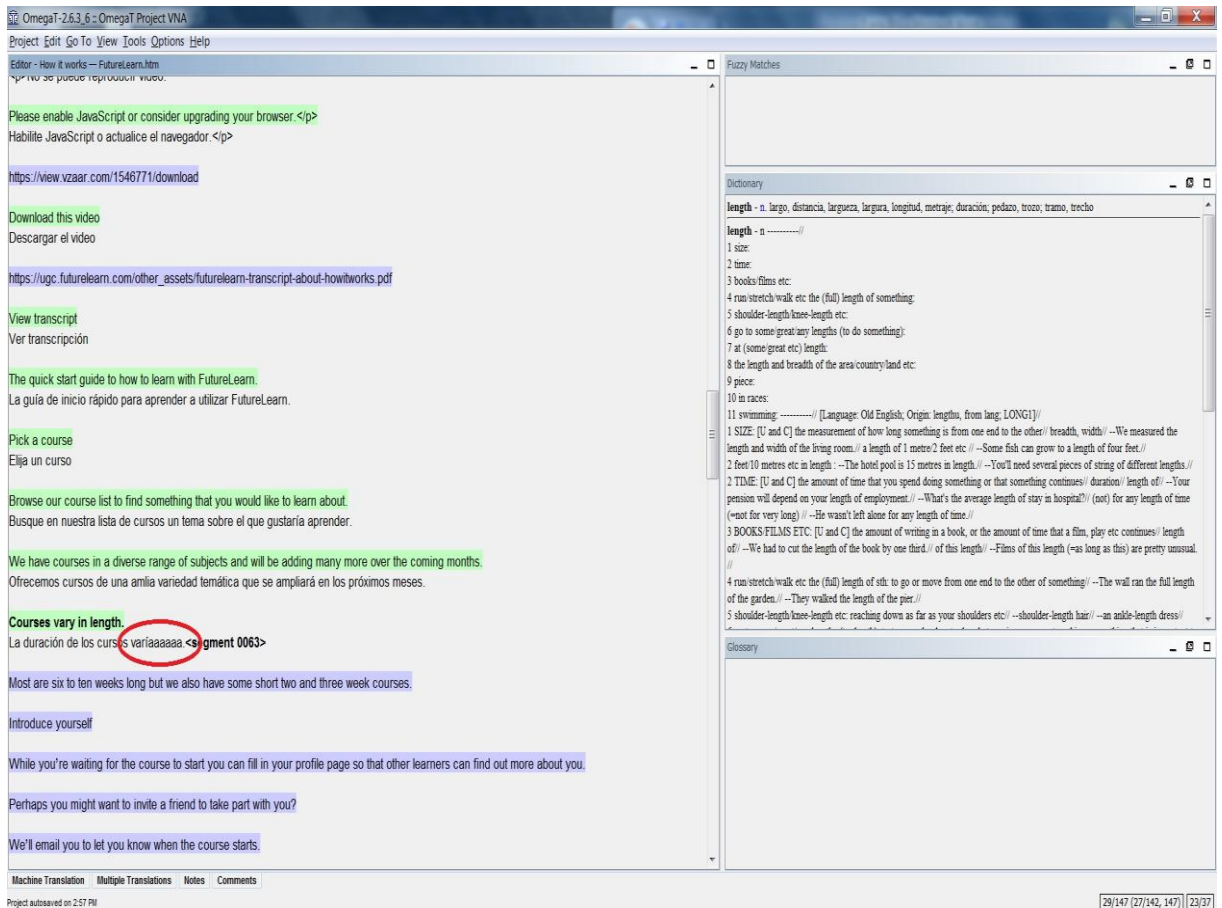


Figure 30. OmegaT editor

There are only some minor differences because of space constraints, as we know that Spanish language usually tends to have a larger extension than English language, which is more contracted. This brings us to a well-known matter, that is, how complex the process of localization is. In localization the translator's job does not end when transferring text from one language into another, but it goes much further trying to adapt cultural and linguistic differences, but always taking into consideration the space restrictions he has on such file format. In addition, as we stated at the beginning of our paper, translators are users of technology rather than developers.

This inconvenience could be partially solved when using OmegaT's *Create Alternative Translation* feature. When we had doubts as regards to the length of a target segments, we used this feature to create multiple translation. When we have retrieved the target file to proofread it in the browser and we have noticed small irregularities as regard to the length, we could always make swift changes in the project, without having to translate again that segment. The following screen shot is an example of such case.

Free CAT tools as an alternative to commercial software: OmegaT

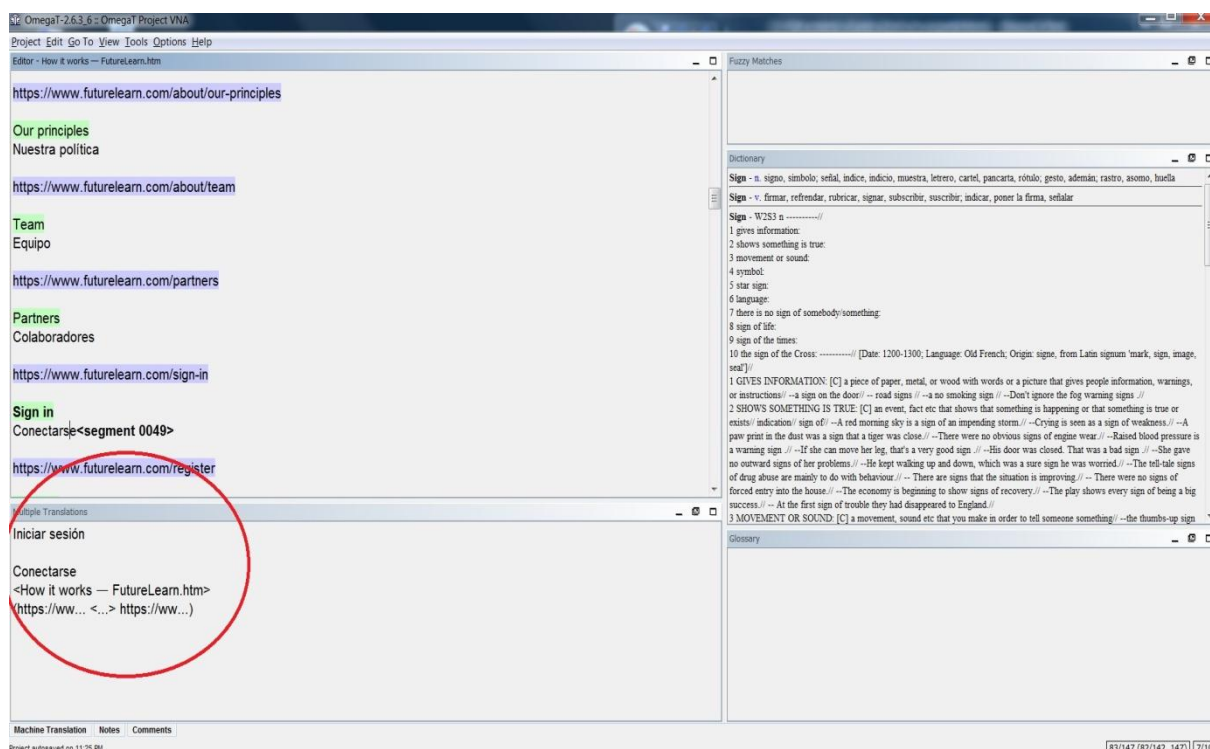


Figure 31. Create alternative translation dialogue window

As regards to our translated webpage, the only visible differences are related with the word extension. For instance, the menu *About* has been translated into Spanish as *¿Quiénes somos?* therefore, the esthetics of the page has been changed. If no other translation equivalences were considered suitable, a solution would be to contact the webpage's administrator to take care of these technical aspects.

In what concerns OmegaT, the translation of the webpage has been smooth, fast and entertaining. The software processes the segments with rapidity, accelerating the translation process. We have been testing its functions and we can strongly affirm that it is a very complete computer-assisted tool which helps increasing productivity and efficiency.

Below we have inserted a screenshot of the OmegaT editor to reinforce our statements on the tool's quality.

Free CAT tools as an alternative to commercial software: OmegaT

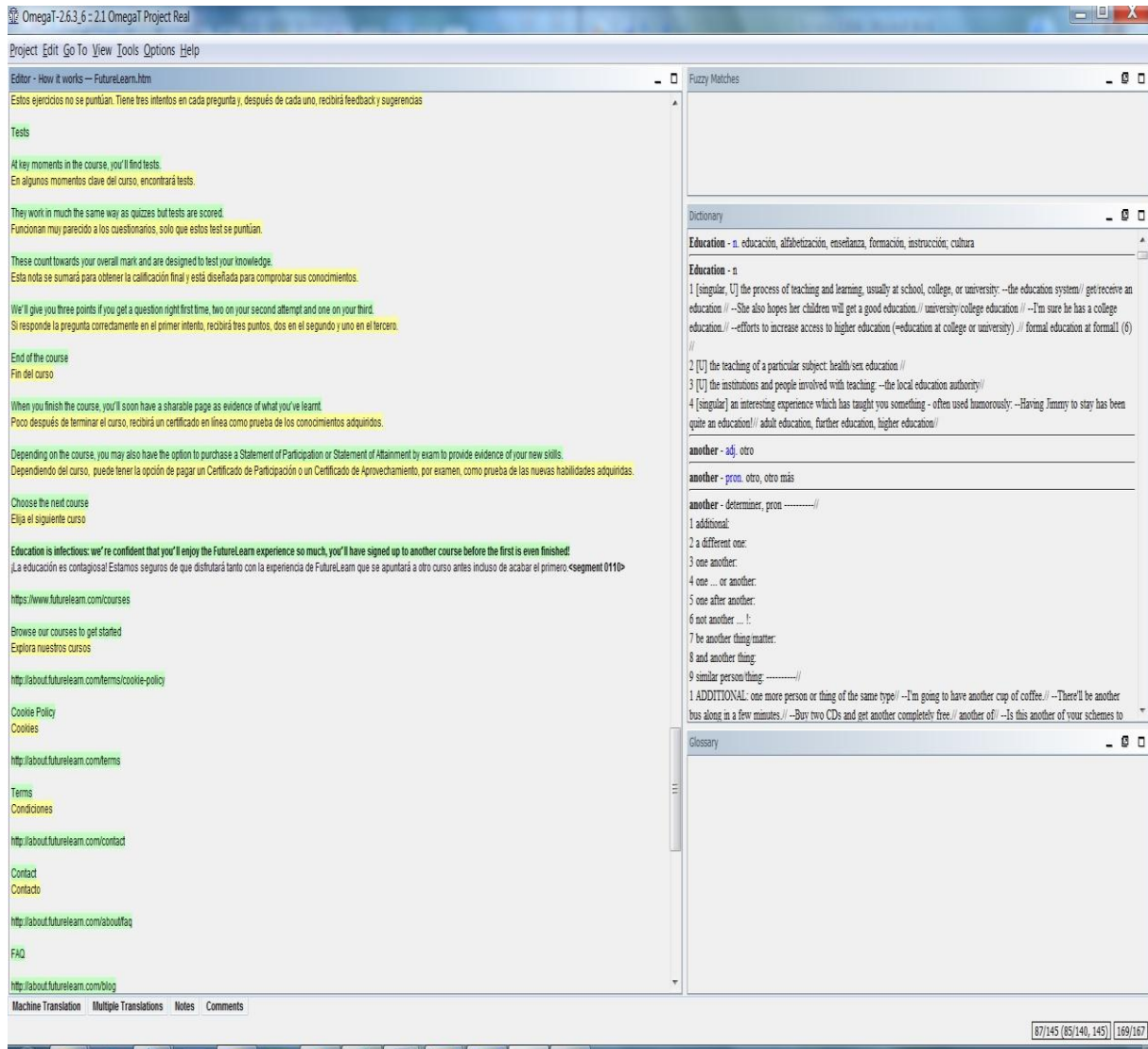


Figure 32. FutureLearn translation process

5. Advantages and disadvantages of OmegaT

The user-friendly interface makes OmegaT an excellent CAT tool for translators, but especially for those who are just beginning to use a CAT tool. One of the main reasons for which some translators are reticent in using a CAT tool is because CAT tools are complex software. In fact, some of them give up as soon as they start because CAT tools may seem complicated if we do not take sufficient time to be familiarized with them. From this point of view, OmegaT is highly commendable because this process is much faster than with other CAT tool.

To learn how to use OmegaT, we can go through the user's guide (there is also a summary available) or simply search the Internet for tutorials or tips from considerate users who have shared their experiences. As regards to this, we must highlight another great advantage of OmegaT, that is, the user support, which is provided for free by OmegaT users. Most commercial CAT tools require an additional cost for offering product support.

In addition, to upgrade to a newer version of OmegaT is free, whilst for a commercial tool upgrade we need to pay a considerable amount of money - depending on the CAT tool.

OmegaT allows its users to install the new version in a new folder and keep the older one, or just overwrite the older version.

Moreover, OmegaT supports a wide range of file formats and, besides XML, TBX it is also supported. Nevertheless, in case we need to convert a file from a format that is not compatible with OmegaT, we can find available on OmegaT's webpage a list with auxiliary tools or we can use OmegaT's plugins. Additionally, we can also make use of plugins to personalize and improve OmegaT, also available for free. As regards to the most common used formats, .docx, PDF and PO files, OmegaT is compatible with all of them. For instance, it manages OCRed PDF files, but with the scripting plugin, more precisely, the scripting language XSLT (Extensible Stylesheet Language Transformation) we can convert XML data into PDF files or even web pages.

The possibility of alternative translations for a segment makes OmegaT an excellent tool, especially in localization when dealing with texts that have space constraints. The *Multiple Translations pane* displays the default translation, the one we consider it would probably suits best, and the alternative one, in case we need it, so we always have a backup.

One of the greatest advantages of OmegaT is that it is a free complete stand-alone CAT tool. Although many translators work from home, to set up a favorable environment and well-equipped it could be very expensive if we have to add the software they need for translating. This is why OmegaT comes in useful; it saves time and expenses.

As regards to time savings, the use of translation memories and fuzzy matches for registering repeated content for future re-use is one of the major factors in increasing productivity

Free CAT tools as an alternative to commercial software: OmegaT

and, therefore, revenues. OmegaT allows us to store translations in a TM and, for future similar projects; we do not need to translate similar segments twice. We can create TMs, as well as terminology databases, for a specific client and keep them confidential or we have the possibility to share them. Additionally, OmegaT allows the simultaneous use of multiple translation memories, if so are the requirements for a specific translation project i.e. a client provides us their own TMs. This is very advantageous when dealing with internal repetitions within texts which require the reuse of entire identical segments i.e. notices, warning, etc.

On the other hand, with the alignment feature, we can reuse previous translations with their original source texts and format them into TMs to be used for similar projects in the future.

Additionally, OmegaT has powerful search functions in source text, in translations, in notes, and in TMs within the current project or in any folder. From the menu *Edit > Search Project* or using CTRL + F shortcut we can open the searching dialogue window. This function is very practical because we can use wild characters such as * and ? depending on the type of search we want to perform. If we want exactly one character match, we use the interrogation mark, but if we want zero or more characters matches to be displayed we type in search box the desired term and the symbol *, as in the example below.

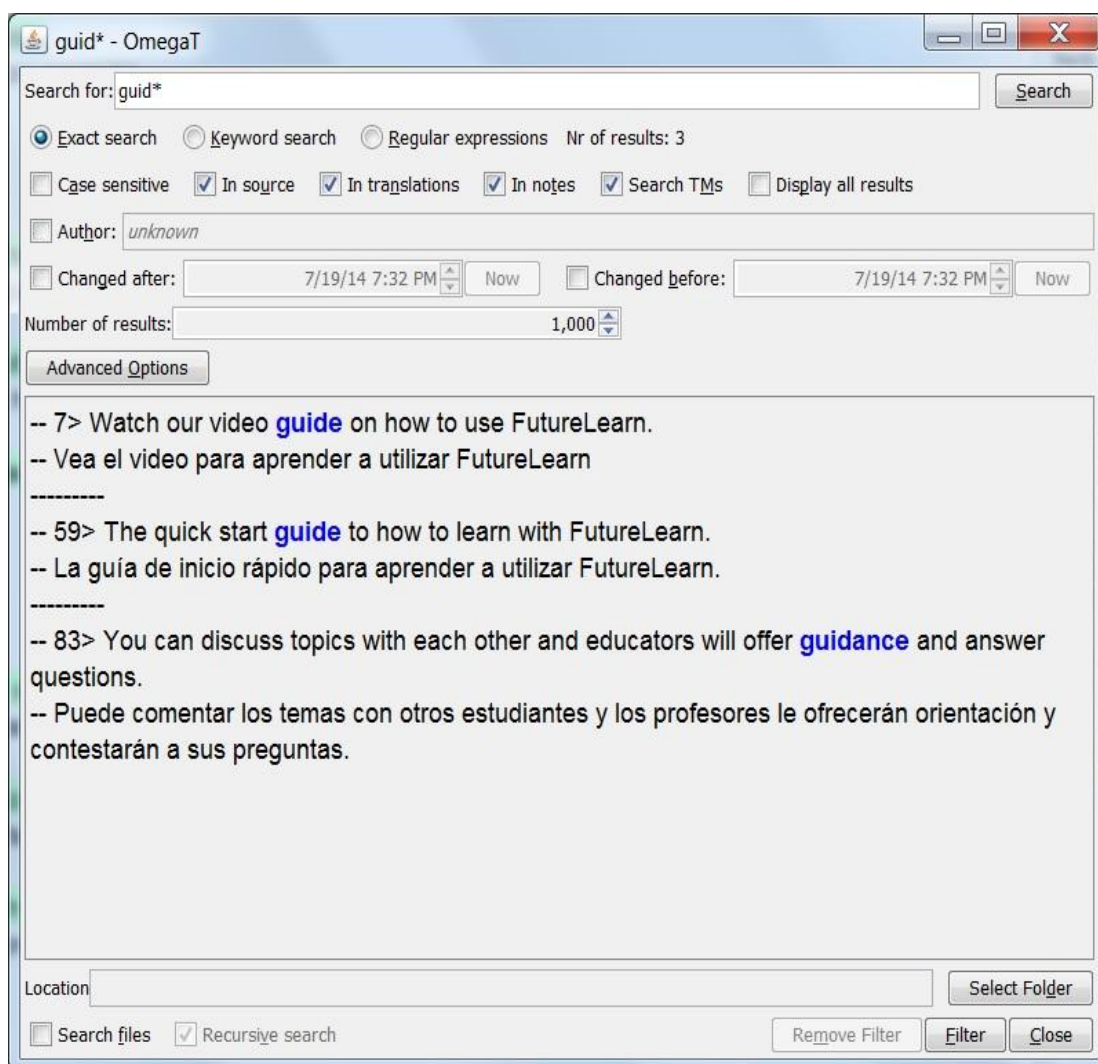


Figure 33. Search dialogue window

Moreover, this search function is very complete allowing us to perform exact searches, keywords searches or simply search for regular expressions. In our opinion, this feature could represent a great strength of OmegaT.

The glossary function not only saves time in terms of switching from OmegaT to a browser or other documentation tools to perform terminology searches, but it also increases the quality of our translations. We can build dictionaries with specific terminology for a certain client, ensuring, therefore, coherence and consistency within translations or throughout a project that requires regular updates.

Although the glossary function in OmegaT could be highly improved, it satisfies the basic needs of a translator in terms of terminology. It does not have too many options of personalizing it and, from what we have experienced, it would have been very handy to use a shortcut to introduce a term from the glossary, but instead, we have to right click on the term in the glossary pane to insert it. It does not represent a great impediment because we do not leave the translation environment, but it supposes that we move from the editor pane to the glossary pane.

Additionally, we can create client-specific terminology, which represents a great strength of OmegaT, or we can import the ones that clients might provide.

All in all, the terminology feature is very easy to use, although it might not be very appealing nor esthetical when managing it.

A small inconvenience we have found on OmegaT is that it does not allow us to split or merge segments, unless we do it through the segmentation rule, which could result somewhat difficult, especially for non-advanced users.

In addition, we have discovered that it does not notify us when misspelling a word, but in exchange, we can run the spellchecker at the end of our work. Besides, the proofreading task depends on the preference of the translator. Some translators prefer proofreading directly in the tool's interface, and some prefer doing this task at the very end, in the target document, for style matters. Nonetheless, in OmegaT segments can be based on paragraphs which can facilitate the proofreading in the interface, if so is preferred.

Its compatibility with operating systems such as Windows, Mac OS, and Linux makes OmegaT a powerful tool suitable for learning and professional environments.

To conclude, it is necessary to precise that OmegaT has more advantages than disadvantages, which makes it a powerful free and open-source CAT tool that can help us increase output, quality and consistency specifying that the benefits of this tool may vary according to the job we are handling.

IV. CONCLUSIONS

Technology is nowadays part of our lives at a level below our notice. Transforming paper-based documents into digital information has represented an enormous step forward for the translation practice. Although paper-based dictionaries, for example, are not completely obsolete, translators have the possibility now to consult not only online dictionaries, but also online data and important sources, which physical distance could have made it impossible. It is certain that now they had to acquire new skills in order to identify which data is accurate and which is not, but this is only a small concession in exchange to the great advantages this technological era has brought.

While we were still debating on advantages and disadvantages of technology, this transition has been already made and we were scarcely aware of. The same happened with the incorporation of CAT tools in the translation practice, and the same is happening with free CAT tools. We just need to be open and embrace change, as we will lose nothing in the transition.

Computer-aided translation tools are, without any doubt, indispensable considering the growing demand for translation in a highly technologically developed society. More and more translators incorporate these tools in their daily *modus operandi* in order to increase their productivity. The preeminent disadvantage correlated with this software is the high price. Nevertheless, it represents a very good investment, but unnecessary as there are free CAT tools on the market. As we have tried to explain and exemplify in our paper with our translation exercise in OmegaT the differences between commercial CAT tools and free ones are not conspicuous because they use almost the same algorithm.

We are constantly using free and open-source software (most of the time without even being aware of this fact) as there is a large range to choose from and we do not apperceive the necessity of complementary features when performing tasks. In the same way, we should eliminate the *restrictions* that come with commercial CAT tools in order to facilitate the access to knowledge and long-life learning. With fully access to software we also promote autonomous learning and independent problem-solving.

All in all, we have to assert that OmegaT is an excellent free CAT tool, which includes many useful advanced features for facilitating the translation process.

We have seen that it is efficient and enjoyable and it really helped us increase productivity and efficiency as our translation exercise can prove. Also, it is responsive and easy to use and it manages very good HTML files.

Apart from the documentation presented in this paper, in support of our account, to highlight the fact that free CAT tools, and more precisely, OmegaT, are very useful and they are

gaining popularity among professionals, we incorporate a testimonial of an English to Spanish freelance translator on a forum in Proz, business-related social networking site for translation, because we would like to emphasize how OmegaT is starting to replace popular, commercial CAT tools.

Here is my specific problem:

I am an advanced user of Trados Studio, and I am trying to make my way out of it and into OmegaT. Why? Well, there is not a single reason...

I am a reasonably proficient Linux user (Ubuntu) and I really hate to have to run a virtual Windows machine ONLY to be able to use Trados. I have been using Trados since its 2007 version, and it gets bulkier and slower with every new version coming out. I won't even comment on Trados price strategy... I could go on, but it's not the point of my question.

So far, I am just learning the basics of omegaT, and I think it could be a good replacement for Trados. One thing I got used to while using Trados is the capability of choosing the language pair (hence the TM) when you start translating a file or project. Is there any way to do something similar in omegaT? (ProZ, 2014)

With our paper, we have tried to prove that OmegaT, which is not only a free tool, but open source as well, is a professional useful software and it should be introduced not only in translators' workstation, but also in Translation and Interpreting students syllabus.

To conclude, we can affirm that the results of the translation exercise were satisfactory and have provided us with enough arguments to convince us to continue using it. We hope that this paper will help other translators and future translators to decide to give it an opportunity.

BIBLIOGRAPHY

- Alcina, A. (2008). Translation technologies scope, tools and resources. *Target*, 20(1), pp.79-102.
- Anastasiou, D. and Schäler, R. (2010). Translating Vital Information: Localisation, Internationalisation, and Globalisation. *Syn-thèses Journal*, 3, pp.11-25.
- Austermühl, F. (2001). *Electronic tools for translators*. 1st ed. Manchester: St. Jerome.
- Bowker, L. (2002). *Computer-aided translation technology: a practical introduction*. Ottawa: University of Ottawa Press.
- Catford, J. C. (1965). *A linguistic theory of translation; an essay in applied linguistics*. London: Oxford University Press.
- Cánovas, M. and Samson, R. (2011). Open source software in translator training. *Revista Tradumàtica: tecnologies de la traducció*, (9), pp.46-56. [online] Available at: <http://revistes.uab.cat/tradumatica/article/view/9/pdf> [Accessed: 18 May. 2014].
- Díaz Fouces., and García González, M. (2008). *Traducir (con) software libre*. 1st ed. Granada: Comares.
- Esselink, B. (2000). *A Practical guide to localization*. 1st ed. Amsterdam: John Benjamins Pub. Co.
- European Union, (2012). *Studies on translation and multilingualism: Crowdsourcing translation*. 1st ed. Luxembourg: Publications Office of the European Union, pp.23-25.
- Fsf.org, (2004). Free Software Foundation: working together for free software. [online] Available at: <http://www.fsf.org> [Accessed: 12 Jun. 2014].
- Gambier, Y. and Doorslaer, L. (2010). *Handbook of translation studies*. 1st ed. Amsterdam: John Benjamins Pub. Co.
- Gurevych, I. & Kim, J. (2013). *The people's web meets NLP*. 1st ed. Berlin: Springer.
- Heer J. & Bostock M. (2010). *Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design*. ACM Human Factors in Computing Systems (CHI).
- OmegaT support (2014). Importing a translation memory into OmegaT. *ProZ.com*. [online] Available at: http://www.proz.com/forum/omegat_support/146666-importing_a_translation_memory_into_omegat.html [Accessed: 1 Jul. 2014]
- Kenny, D. (1999). CAT Tools in an Academic Environment: What Are They Good For?. *Target*, 11(1), pp.65-82.

Lawlor, B. (2007). *The Age of Globalization: Impact of Information Technology on Global Business Strategies*. The Honors Program Senior Capstone Project. Bryant University.

Lingo Systems & ATA. *The guide to translation and localization: preparing for the global marketplace*. (2002). Portland. ATA

McKay, C. (2006, June). Free and Open Source Software for translators. *Panacea*, VII, 95-98. [online] Available at: <http://www.medtrad.org/panacea/IndiceGeneral/n23_tribuna_McKay.pdf> [Accessed: 23 May 2014].

O'Hagan, Minako and Ashworth, David (2002). *Translation-mediated Communication in a Digital World*. Clevedon: Multilingual Matters Ltd.

Pym, A. (2004). *The moving text localization, translation, and distribution*. Amsterdam: John Benjamins Pub. Co.

Quah, C. K. (2006). *Translation and technology*. Houndmills. England: Palgrave Macmillan.

Reinke, U. (2013). State of the Art in Translation Memory Technology. *Translation: Computation, Corpora, Cognition*, 3(1).

Somers, H. L. (2003). *Computers and translation: a translator's guide*. Amsterdam: J. Benjamins.

Smolej, V. (2013). *User's Guide*. Unknown: Vito Smolej.

Sourceforge.net, (2014). *SourceForge - Download, Develop and Publish Free Open Source Software*. [online] Available at: <<http://sourceforge.net/>> [Accessed: 27 Jun. 2014].

"The Open Source Initiative." *The Open Source Initiative*. [online] Available at: <<http://opensource.org/>> [Accessed: 19 Jun. 2014]

TranslateMedia, (2014). *Going Freelance - Guide to Becoming a Freelance Translator*. [online] Available at: <<http://www.translatemediamedia.com/careers/how-to-become-a-translator/going-freelance/>> [Accessed: 25 Aug. 2014]

Unesco.org, (2014). Free and Open Source Software (FOSS) | *United Nations Educational, Scientific and Cultural Organization*. [online] Available at: <<http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/free-and-open-source-software-foss/>> [Accessed: 25 Jun. 2014].

Vallianatou, F. (2005). Cat tools and productivity: Tracking words and hours. *Translation Journal*, 9(4). [online] Available at: <<http://translationjournal.net/journal/34CAT.htm>> [Accessed: 14 Jun. 2014].

ANNEXES

The current paper comprises additional material to reinforce the assertions as regards to free CAT tools, and, to be more precisely, OmegaT.

Annex 1 represents a list from ProZ of the most common CAT tools among translators.

Annex 2 represents a short video explaining how to add dictionaries to an OmegaT project.

Annex 3 encloses a list of screen shots with the original source text and the target text.