

# Usability of Free and Open-Source Tools for Translator Training Omegat and Bitext2tmx

María Teresa Veiga Díaz, Marta García González  
University of Vigo, Spain

## 1 Introduction

In Spanish universities, free and open-source software (FOSS) is widely used in technical areas because of its usability, adaptability and low cost. Conversely, the use of these tools in the field of translator training has been minimal despite the existence of suitable software specifically developed for translation activities, such as OmegaT, Anaphraseus, bitext2tmx, Sun Open Language Tools, ForeignDesk or Transolution. In this context, GETLT was created to promote the use of FOSS both in translator training and professional translation, and to acknowledge the effort made by FOSS localization teams. After a short overview of the phases and results of research project PGIDIT07PX1B302200PR, Creación dunha plataforma docente GNU/Linux para a formación de tradutores – localizadores de software – subtituladores, funded by the Galician Government, within the framework of programme Incite, this chapter describes a particular research effort focused on testing the usability and applicability to translation training of free and open-source translation memory managers and text aligners with different texts types and genres.

### 1.1 The Background Project

The purpose of the initial project was to develop a computer environment for the training of translators and interpreters based on free, open-source software, more particularly a GNU/Linux distribution in a live-CD that could also be installed on the computer's hard disk, to be freely used at translation training university centers worldwide, and adapted to meet the particular needs of the educational programs at each university. The underlying idea was to develop an environment that could be used for translator training in all the different courses comprising a degree in translation and interpreting. It should facilitate the use of CAT tools for translator training, by removing the high costs of proprietary licenses, and also encourage the use of free, open-source software among students, future professional translators, thus covering the existing gaps within this group as concerns free software (Fernández García 2006a: 76-80; García González 2008: 9-31).

The activities in the project were arranged in four different phases, some of which were developed simultaneously rather than on a strict consecutive

basis. Although it is beyond the scope of this paper to discuss in detail each phase and the project's results (García González, 2013), a short description of the activities and main results follows:

**Phase 1: Analysing training requirements in the different varieties of language mediation**, by means of interviews to teachers and translation professionals, and **choosing a series of free software applications** running over GNU/Linux O.S. that were able to meet such requirements. The interviews were carried out both in situ and via e-mail and the information compiled was used as a basis for the subsequent phases of the project.

**Phase 2:** Following the above data (requirements and chosen applications), **generating a GNU/Linux distribution that was both live executable from a live-CD and installable on the computer's hard disk, targeted to the training of language mediation professionals**. The distribution was generated, based on Linux Mint Distribution, under the name of MinTrad (for a detailed description of this and other Linux Distributions for translators, see Sandrini in this same volume).

**Phase 3: Disseminating the project's results** within the university community: Results were presented at several conferences and also described in different papers and chapters during and after the duration of the project.

**Phase 4: Documenting the distribution in a complete and sufficient manner**, by preparing a comprehensive user guide for all the tools and applications comprising the distribution, and **testing the environment** both with students and with professional translators. This phase was planned as a long-term activity, as it could not be fully covered within the duration of the project. A short part of the testing effort is described in this chapter.

## 1.2 Documenting and Testing MinTrad

The distribution prepared under phase 2 of the project, MinTrad, included 30 computer-aided translation applications, among which one text aligner (bitext2tmx), and four translation memory managers (OmegaT, Anaphraseus, Transolution XLIFF editor, Sun Open Language Tool). As already mentioned, in addition to the preparation of the distribution, the project envisaged a phase focused on documenting and testing the applications in terms of their usability both in different types of translation courses and in professional translation situations. Here, usability is understood as the effectiveness, efficiency and satisfaction with which translation trainees and professionals achieve specified translation goals in a formative or professional environment, which is

in agreement with the standard definition of usability (ISO 9241). The satisfaction of translation trainees with the MinTrad distribution was preliminarily measured in previous phases of the project (García González, 2013) through a survey conducted among translation students. The survey included questions on their familiarity with FOSS, the complexity of the distribution, and the usefulness of MinTrad in translator training environments. Overall, fourth-year students, who had the opportunity to test the distribution with different types of texts, showed satisfied with the usefulness of the distribution in didactic settings and considered that it would be even more useful for use in professional environments. Yet, the survey did not include questions on the usability of specific tools or on the effectiveness and efficiency of the distribution. Accordingly, to complete the results of the previous phases of our research, the usability and applicability of two free and open-source computer-assisted translation tools included in the MinTrad distribution, namely OmegaT and bitext2tmx, were tested. The main purposes of the tests were (i) to determine the advantages and drawbacks of the tested applications as compared to similar proprietary software applications; and (ii) to determine the applicability of the translation memories generated by using the tested applications with different types of texts in the specialized translation classroom.

## 2 Materials and Methods

### 2.1 Tools

Two software applications were tested, bitext2tmx text aligner v. 1.0MO and OmegaT translation memory manager, versions OmegaT\_2-2-2\_04\_Beta and OmegaT2.1.7\_02 for Linux. As mentioned in section 1, both applications are free and open-source and are included in the MinTrad distribution. Bitext2tmx and OmegaT were tested under three operating systems, Windows XP, Linux MinTrad and MacOS X, insofar as it was assumed that the possibility of using the applications regardless of the operating system used was a big asset for translator trainees, who are not constrained to use a specific system. Actually, the computers available to our students both in free-access rooms and in classrooms have two partitions, one for Windows and another one for Linux.

Bitext2tmx (<http://bitext2tmx.sourceforge.net/doc/guide/en/Bitext2tmx.html>) was originally developed by members of the Transducens research group at the department of languages and computer systems of the University of Alicante, Spain. As a text aligner, bitext2tmx allows for the creation of translation memories in TMX format by aligning an original text and its translation, both in plain-text format. The generated memories can be edited and aligned to provide better matches when used with any translation memory manager.

The tested text aligner was not further developed, such that no more recent versions are available.

OmegaT (<http://www.omegat.org/en/omegat.html>) is probably the most widespread free cross-platform translation memory application and has been the focus of several papers in the past few years (Carretero 2010; García 2010; Prior 2010). It is intended for professional use and commonly used by translation students at the University of Vigo. Among its features are: fuzzy matching, simultaneous use of multiple translation memories, user glossaries with recognition of inflected forms, more than 30 file formats (including Microsoft Office 2007 and later, PDF, HTML and XHTML, ODF, PO, and IDML/TTX/XLIFF/XML), spelling checker, compatibility with other translation memory applications and interface to Google Translate. It is under constant development and has gradually incorporated new features. The most recent stable version of the application is OmegaT 3.1.9.

## 2.2 Methods

To determine the usability of the selected tools, the three components of usability, namely effectiveness, efficiency and satisfaction (Jordan 1998) were explored. Effectiveness was understood as the accuracy and completeness with which translators can achieve the relevant goals, i.e. a satisfactory alignment of two parallel texts or a satisfactory translation with a high percentage of matches; efficiency was understood as the resources expended in relation to accuracy and completeness in terms of time, money and knowledge required to use the tool and, finally, satisfaction was understood in terms of the comfort and acceptability of the system to the users. The method used to test the usability of the applications and to determine the applicability of the generated translation memories was divided into three phases: i) text alignment and generation of translation memories; ii) application to translation projects and iii) application to learning environments. The effectiveness and efficiency of the tools were analyzed in all three phases, while comfort and acceptability were studied mainly in the first phase of the analysis according to the following four criteria: accessibility and installation, interoperability, functionality and interface.

### **i) Text Alignment and TM Generation:**

In the first phase, the translation memories that would later be fed into the TM manager were generated with `bitext2tmx`. To this end, a parallel text corpus was compiled. Also, a monolingual corpus was compiled to later test the usability of the OmegaT TM manager through the simulation of a number of translation projects. Both corpora included three sub-corpora, a sub-corpus of legal texts, a sub-corpus of economic texts and a sub-corpus of scientific and

technical texts. The selected texts were saved in different file formats, namely \*.doc, \*.txt, \*.odt, \*.rtf, \*.pdf and \*.ppt, such that the usability of both tools could be tested. The scientific sub-corpus was composed of only three genres: scientific papers, patient information leaflets (PILs) and game console user guides. The scientific papers included in the corpus were originally written in Spanish and translated into English, and focused on farm production and classification. The genres covered by the economic sub-corpus included corporate reports, annual accounts, cost and financial accounting reports, SAP user instructions, and press releases, while the legal sub-corpus included testaments, articles of incorporation, agreements, legal forms and EU legislation. Contrarily to scientific papers, the legal and economic texts were in their most part originally written in English and translated into Spanish, except in the case of EU legislation, of which no reference was found to which was the original text of the pair.

In some cases, individual translation memories were created from each pair of texts, but in other cases, as with testaments or corporate documentation (annual reports or UE legislation), the individual translation memories were merged with the help of an OmegaT plug-in, TMX-Merger, a Java command-line script for merging two or more TMX files. A total of 114 pairs of texts of different lengths, ranging 75 to 15800 words were aligned. In this phase, the effectiveness of bitext2tmx was determined by defining the accuracy with which the selected pairs of texts were aligned, and efficiency was determined based on the resources needed to complete the task. As per satisfaction, four criteria were considered: accessibility and installation, interoperability, functionality and interface.

## **ii) Application to Translation Projects:**

The memories generated in the first phase of our research were fed into the projects. A total of 11 translation projects were created, three of which corresponded to scientific and technical texts, another three to economic texts, and the remaining seven to legal texts. From among the seven legal translation projects, five corresponded to texts extracted from the EUR-Lex database and were analyzed as a unit. As in the text alignment phase, the selected source texts had different lengths so that the performance of the tool could be studied separately. For all text types, the texts selected for validation were similar to those used in the specialized translation classroom. In this case, the effectiveness of OmegaT was determined based on the number of 100% and fuzzy matches, and efficiency was analyzed in terms of the time and effort required to achieve an accurate and complete translation using the TM fed into the project. As in the first phase, the satisfaction of users was determined based on accessibility and installation, interoperability, functionality and interface.

### iii) Application to Learning Environments:

After the texts were aligned and the performance of the generated TM was tested in OmegaT, the last phase of the project consisted in testing the tools in a specialized translation classroom, particularly in a scientific and technical translation course of the fourth year of the Degree in Translation and Interpreting. Three translation projects were created, one for each of the selected genres, a specialized paper, a PIL and a game console user's manual. The purpose of the test was to try both tools with the most common types of texts in the classroom and assess their benefits and drawbacks for translation trainees. Thus, students would learn: (i) to determine when and with which resources it is effective and efficient to use CAT tools; (ii) to identify the factors that affect the quality of a translation performed with these tools; (iii) to assess the suitability of the machine translation solutions provided by the TM manager. The criteria used to assess the usability and applicability of the tools in this phase were the same as in the second phase of the project, but the formative nature of the translation projects was considered.

## 3 Results

In this section, we present the results for the three phases of the project. First, we provide an overall assessment of the performance of bitext2tmx and OmegaT (for a thorough discussion of the quality of the translation memory manager, please see Flórez & Alcina in this same volume). Then we focus on the results of the application of both tools to particular translation projects, both professional and formative, for the three types of texts considered, business, legal and scientific, and technical.

### 3.1 Overall Assessment

#### 3.1.1 Bitext2tmx

The main benefits of the text aligner included in MinTrad are related with accessibility and ease of use and installation, whereas the main drawbacks are related with efficiency. Bitext2tmx is a free and open source text aligner that requires no installation. It runs on the three operating systems tested, Windows, Mac and Linux, and generates .tmx files that are compatible with other CAT tools, both free and proprietary.

In didactic settings, bitext2tmx is highly applicable, because it is intuitive and easy to use for beginners. In addition, it runs smoothly with short, edited texts and the results for these texts are good, which makes it particularly suitable for use during the first years of the degree, when students start using CAT tools and translating very simple texts.

Despite these benefits, bitext2tmx has a number of problems related to its functionality that make it less efficient for use among advanced users or with longer texts than similar proprietary tools. Particularly, the following drawbacks have been observed during our testing:

Although the application runs on the three operating systems, it does not recognize files with hidden extensions in Mac OS X. Moreover, only \*.txt files can be aligned, such that other types of files must be converted before alignment, which requires spending more time and effort.

Bitext2tmx does not allow for saving partial alignments, which can be seriously inconvenient when working with long texts. In addition, changes are not saved in case of a shutdown of the application, such that the users need to start over again, thus losing efficiency. Furthermore, alignment of more than one pair of texts per project is not enabled. Therefore, users cannot generate a single translation memory (TM) for several texts and each generated translation memory corresponds to a single pair of texts, thus forcing the use of a TMX merger. In bitext2tmx, alignment rules do not seem to consider language pair specificity, such as the average sentence length or the presence of graphical accents, which requires pre- or post-editing by the user in order to obtain a reliable TM. Moreover, some symbols and signs, such as those for percentages, decimals, semi-colons, among others, are often misinterpreted as full stops, which seriously affects segmentation and, therefore, effectiveness.

Finally, the application is not as user-friendly as similar proprietary tools because the interface lacks some functionality such as keyboard shortcuts, the scroll function for the translated-text window, or mechanisms for simultaneous selection of several lines of text. Yet, the “split by line break” functionality partially improves segmentation, particularly for tables and figures.

The above assessment suggests that bitext2tmx is a simple tool that can be useful for students who are involved with the translation of short, simple texts, but not for professional translators who prioritize efficiency.

### **3.1.2 OmegaT**

OmegaT is an easy-to-use-and-install tool that runs on the three operating systems, although it requires reading the manual for the creation of new projects. In addition, OmegaT does not support every file extension, \*.txt, \*.docx and \*.odt files are supported, but \*.doc files are not supported. Yet, the main drawbacks of this free and open source application are related to its functionality.

As regards segmentation, OmegaT segments into paragraphs, with no segment expansion or shrinking enabled on the interface. If sentence segmenta-

tion is preferred, the text must be pre-edited and rules must be setup in the main menu, in Options → Segmentation. In addition, the application does not correctly identify the matches with long paragraphs, such that both effectiveness and efficiency are affected.

With regard to terminological extraction features, the application enables the generation of glossaries, but glossary terms cannot be automatically extracted, such that terms must be manually added to the project glossary. In addition, the glossary is necessary to retrieve specific terms because the application does not find matches by term. Yet, generating glossaries in OmegaT is very simple, insofar as glossaries are lists of words separated by a tab. In didactic settings, this is an advantage insofar as it allows students to reuse the glossaries prepared for every course and feed them into any project. In contrast, TMs from other projects or translators that have been generated with tools different from OmegaT, such as the bitext2tmx aligner, can be used as ancilliary translation memories but not directly imported into the master translation memory of the project, `project_save.tmx`, unless merged through the `TMXMerger` java command-line script. Working with many ancilliary TMs may unnecessarily slow OmegaT down, thus reducing the efficiency of the tool. In addition, ancilliary translation memories are read by OmegaT but not corrected during the project, which reduces the efficacy of the tool. Therefore, merging the TMs from other translation or alignment projects with the master TM speeds up the process and makes it more reliable. Nevertheless, merging `.tmx` files with `TMXMerger` requires some level of programming and might be tricky for some students, particularly for those who do not have specific training.

Another problem related with TM creation is the fact that wrong translations are not deleted when corrected unless they are stored in the main TM, which can affect the accuracy with which the relevant task is performed. Other efficiency issues are related to the creation of labels; OmegaT inserts “fuzzy match” labels that are not automatically removed when the final files are generated, such that users must remove these labels every time that an insertion is confirmed or when the final file is generated.

It should be noted that as versions OmegaT 2-2-2\_04\_Beta and OmegaT 2.1.7\_02 for Linux were used in the test, some of the drawbacks referred to above might have been already corrected in later versions. In addition, despite the drawbacks, which can be rather limitative to professional users, OmegaT has many benefits for use by students. First, OmegaT is a free and open source tool that runs on the three OS tested and is already installed in MinTrad. The application includes a complete and relatively simple user manual and a readily accessible quick start guide that is very useful for



students who are starting to become acquainted with the application. The translation process is simple and intuitive, in contrast to project creation, which requires reading the manual. Once the project is created, the application is easy to use and the interface is user-friendly: it enables keyboard shortcuts, which speeds up the process, and incorporates machine translation options (Google Translate, Apertium, Belazar). The possibility to search Google Translate can be useful sometimes, but it must be handled with care in didactic settings, in order to avoid random use of the option by students.

Also, OmegaT retrieves up to five matches, indicating percent match and origin, which is useful when different unmerged TMs are used. In addition, the application allows alternative use of various files within the same project. Finally, the application offers some utilities, such as a text aligner and a tmx merger. Yet, as explained above, using these utilities requires specific knowledge of java script, which makes it complex for inexperienced students.

In the following sections, the results of the applicability of the generated TMs for the translation of each text type and genre are discussed.

## **3.2 Applicability to Translation Projects**

According to the test results, the applicability of the text aligner and the generated TMs depends strongly on text type and genre.

### **3.2.1 Business Texts**

- Financial reports: good results were obtained both with TM manager and aligner when translating reports from the same company for different years. Otherwise, results were poor except for audit reports.
- General meeting agenda: again, results were highly satisfactory when the TM manager was used for the translation of agendas from the same company. When translating texts from other companies, though, results were poor except for legal fragments connected to companies law.
- SAP training presentations: several problems were encountered during the alignment of the (ppt) presentation, mainly connected to the conversion of text for alignment. However, after editing, the TM proved rather effective with similar SAP Training Documents.

### **3.2.2 Legal Texts**

- EUR-Lex legal texts: overall, the use of the translation memories resulting from alignment of EU legal texts proved highly effective for the translation not only of other EU texts but also of acts from the different Member States that were adapted to EU law.

- Articles of incorporation: as in the case of financial reports, results with articles of incorporation were satisfactory when translating document amendments but rather poor with texts from different companies, except for legal fragments connected to companies law.
- Service agreements: although results were excellent with short texts, particularly with agreement forms, longer texts produced fewer match retrievals, particularly in sections containing specifications, which decreases the effectiveness of the tool.

### 3.2.3 Scientific and Technical Texts

- Specialized scientific papers: overall, the applicability of the generated TMs to scientific papers is very limited. Actually, the TMs generated from the text pairs used to test the aligner were useful only for papers with a high percentage of complete paragraphs repeated from previous papers. Accordingly, the usability of the tested tools for this text genre is very poor.
- User manuals of simple electronic devices: in contrast to the results obtained for specialized papers, both bitext2tmx and OmegaT showed highly usable for the translation of user manuals of different versions of simple electronic devices, provided that the quality of the aligned texts was good.
- Product information leaflets (PILs): the applicability of the generated TMs was excellent, in terms of both effectiveness and efficiency. Some comfort issues were observed, but the overall performance of the tool with this type of texts was very good.

## 3.3 Applicability to Translator Training Environments

To test the applicability of the tools to formative translation projects, the students of the course in Scientific and Technical Translation at the University of Vigo were asked to create three translation projects in OmegaT using the TMs generated in the first phase of our research, as mentioned earlier in this paper. In this section, the results of the activity are discussed.

- *Specialized scientific papers*: The results for effectiveness were very poor for this genre because of the extremely low percentage of 100% or fuzzy matches obtained by students. Actually, the TMs generated from the text pairs used to test the aligner were almost useless for the translation project tested in the classroom because of the low percentage of complete paragraphs repeated from previous papers. The number of matches retrieved with the tool was so low that it was highly inefficient. Efficiency could increase if the terminological management utility was improved, particularly to guarantee terminological consistency among papers by the same

authors. In addition, the solutions provided by Google Translate in this case were almost useless. Yet, the activity helped students learn to handle machine translation with care because of the evidently poor automatic translations retrieved. Consequently, despite the poor results, this type of project is useful as a formative tool for students insofar as they learn through practice that the applicability of the generated TMs for scientific papers is very limited.

- *User manuals of simple electronic devices*: very good results in terms of effectiveness and efficiency were obtained with instructive texts that corresponded to user manuals of different versions of the same game console. Provided that the selected texts correspond to simple devices, which are usually short, this type of text is highly applicable in the translation classroom for students who are not well-acquainted with text aligners and CAT tools. Yet, the quality of the translations strongly depends on the quality of the aligned texts. Therefore, the quality of the aligned translated texts will determine the teacher decision on whether it is efficient to use a text aligner to generate a translation memory. Alternatively, a good translation memory can be generated from the translation of short texts that are revised and corrected in the classroom, instead of generating a memory from translations available from the internet, as was the case of one of the texts tested in this phase of the project (see Figure 1).

290	To remove your game, first turn the power off.	Para remover el juego, primero apague el sistema.
291	Push the DS Game Card into Slot 1 until it clicks.	Empuje la Tarjeta de Juego de Nintendo DS hacia adentro de l...
292	It will automatically eject partway out of the slot.	De forma automática la tarjeta será expulsada parcialmente d...
293	Using Game Boy Advance Game Paks 1.	Uso de Cartuchos de Juego de Game Boy Advance 1.
294	Make sure that the Nintendo DS is turned off.	Asegúrese de que el Nintendo DS esté apagado.
295	2.	2.
296	Insert the Game Boy Advance Game Pak into Slot 2 on the front...	Inserte el Cartucho de Juego de Game Boy Advance en la ranur...
297	Make sure it is fully inserted into Slot 2.	Asegúrese de que el cartucho esté insertado por completo en l...
298	The label should face towards the bottom of the DS.	La etiqueta debería mirar hacia la base del Nintendo DS.
299	(Illustration 7)	(Ilustración 7)
300	3.	3.
301	Turn the power ON.	Encienda el equipo.
302	The game title will appear.	El título del juego aparecerá.
303	Touch the game title with the stylus to start the game.	Toque el título del juego con el stylus para empezar el juego.
304	Refer to the instruction manual for the game you are playing fo...	Refiérase al folleto de instrucciones para el juego que esté jug...
305	4.	4.
306	To remove a Game Pak, first turn the power off.	Para remover el juego, primero apague el sistema.
307	Push the cartridge out of Slot 2 with your thumb When you are f...	Usando el dedo pulgar como se indica, empuje el cartucho hac...
308	Do not wrap the AC Adapter cord around the DS.	Al dejar de jugar siempre apague su sistema y desconecte el...
309	When not in use, close the DS to protect the screens from dust...	No envuelva el alambre del Adaptador de Corriente en el Ninte...
310	NOTE:	Cierre el DS cuando no se encuentre en uso para proteger a la...
311	The power will NOT automatically turn off when the system is cl	NOTA:

Figure 1: Alignment of an original text and a poor translation that makes the use of *bitext2tmx* inefficient.

- Product Information Leaflets (PILs): the performance of the text aligner and the TM manager was good for this genre. The stability of the macrostructure and phraseology of this genre makes it suitable for testing both

effectiveness and efficiency. A single pair of texts was aligned by students and fed into the project as a \*.tmx file. Then, students were asked to translate the PILs for other presentations of the same drug, commercialized in Great Britain and Ireland with different names. A total of three PILs were translated using OmegaT but the process could be successfully extended to the PILs of every presentation of the same product. The results were excellent, and a total of 266 exact matches were found, which accounts for over 95% of the text (see Figure 2).

Editor — Cipralelex-Drops-10mg.txt

GB-107-06-10-13406 FA-NL  
GB-107-06-10-13406 FA-NL<segmento 0001>

PROSPECTO:

Cipralelex 10 mg/ml

Gotas orales en solución

Lea todo el prospecto detenidamente antes de empezar a tomar el medicamento

- Keep this leaflet.
- Conserve este prospecto, ya que puede tener qu
- Si tiene alguna duda, consulte a su médico o farm
- This medicine has been prescribed for you.

Do not pass it on to others.

- Este medicamento se le ha recetado a usted y no
- tengan los mismos síntomas, ya que puede perjud
- Si considera que alguno de los efectos adversos
- efecto adverso no mencionado en este prospecto,

[parcial]Contenido del prospecto:

- 1.
1. Qué es Cipralelex y para qué se utiliza.
- 2.
2. Antes de tomar Cipralelex.

Estadística de coincidencias

	Segmentos	Palabras
Repeticiones:	4	4
Coincidencias exactas:	266	2980
95%-100%:	0	0
85%-94%:	0	0
75%-84%:	0	0
50%-74%:	7	32
No hay coincidencias:	1	4

Figure 2: Almost automatic translation of a PIL using OmegaT.

PILs are commonly used in general and scientific translation courses and provide translation teachers with an excellent opportunity to successfully use free and open source CAT tools in the classroom. One of the benefits of using this genre is that text alignment is highly effective because of the fixed macrostructure and the length of the texts involved, which render the translation of similar texts efficient and effective. Yet, some drawbacks related

to satisfaction were observed by students. First, the text aligner and the TM manager segmented texts differently, such that post-edition was required after translation to avoid the presence of untranslated segments or format issues.

Second, when segments were not identical, the application did not recognize identical matches for some portions of text, such that the suggestions made by the application were not correctly prioritized (see Figure 3) and the suggested partial match was poorer than other available partial matches.

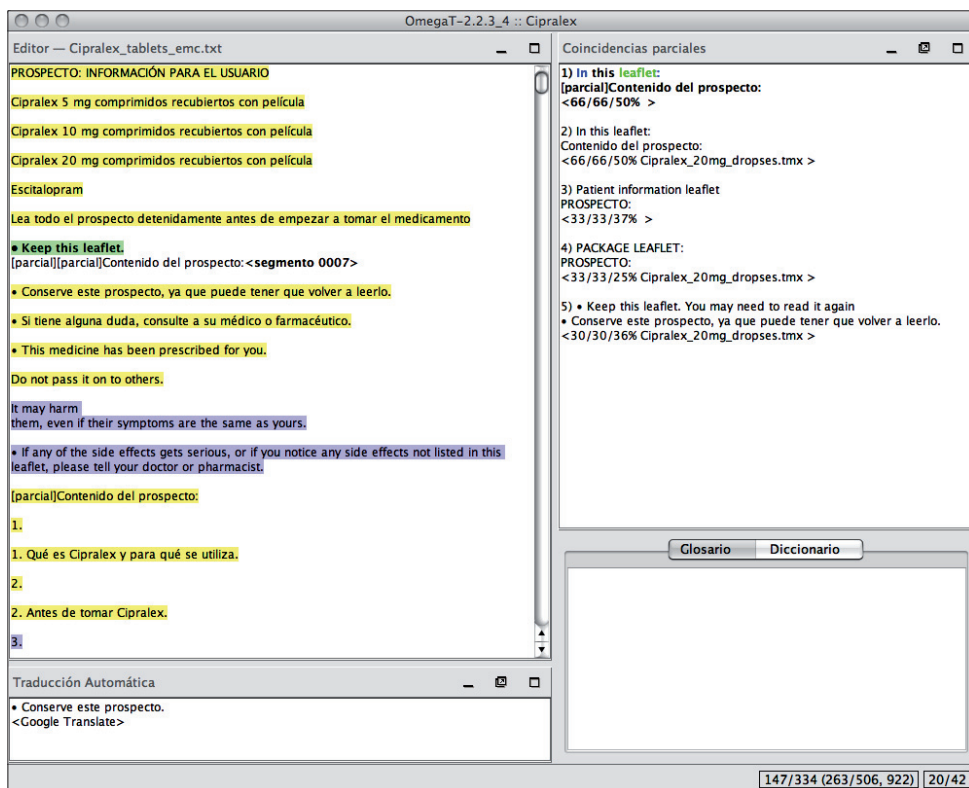


Figure 3: Wrong prioritization of partial matches due to rigid segmentation rules.

## 4 Conclusions

As revealed by the results of the implemented translation projects, OmegaT performs much better than bitext2tmx in terms of effectiveness and efficiency, but the text aligner is easier to use, which increases the satisfaction of users. Overall, the usability of both bitext2tmx and OmegaT seems to be poorer than

the usability of similar proprietary software applications, but they can be used in translator training environments for a number of reasons.

First, `bitext2tmx` allows for the generation of TMs without the need to translate a large number of texts before generating a large TM that can be effective, thus reducing the time required to build useful translation memories from the texts translated in the classroom. Yet, there must be a balance between the time devoted to alignment and the time devoted to translation insofar as text alignment becomes inefficient if the percent of matches is low. Alternatively, students could use TMs available from the Internet. Yet, using this type of resources could be detrimental to students who are not well-acquainted with translation strategies.

Second, `bitext2tmx` helps students better understand how CAT tools work. When using an alignment tool first and then combining the resulting TM with a TM manager, students become aware of the manner in which texts are segmented and may check if this segmentation is appropriate for correct translation. This turns alignment into a relevant learning activity in the first phases of a translator training program.

Finally, OmegaT brings students closer to professional translation environments, in which productivity criteria prevail. On the other hand, using the tool with different types of texts enables them to determine its level of usefulness in different translation contexts. Particularly, they can realize that within the same course, a TM manager is highly productive for the translation of some genres and totally unproductive when translating other genres. Eventually, by using CAT tools and identifying their benefits and drawbacks, students realize that these tools are just tools, and not translators and that it is critical that they are competent translators before they can make the best of TM managers.

In sum, because the professional translation market increasingly demands the use of this type of tools, the translators-to-be need to have knowledge of the performance of the tools, not only of their benefits but also of their drawbacks. For this reason, `bitext2tmx` and OmegaT can be used as a “starter” in training students in the use of CAT tools despite the drawbacks observed during testing and reported in this paper.

## References

- Carretero, I. (2010) Free Software and Translation: OmegaT, a free software alternative for professional translation. In Boéri, J. and Maier, C. (eds.) *Compromiso social y Traducción / Interpretación*. Granada: ECOS.
- Fernández García, J. R. (2006a) La traducción del software libre. Oportunidad de colaborar. *LINUX Magazine*, 19, 76-80. Available at: <http://www.linux-magazine.es/issue/19/Educacion.pdf> [Accessed: 28 January 2013].

- García, E. (2010) Itzulpenak egiteko kode irekiko eta doako laguntzak. In *Senez* 43, 213-218.
- García Gonzalez, M. (2008) Free Software for translators: is the market ready for a change? In Diaz-Fouces, O. and García González, M. (eds.) *Traducir (con) Software libre*. Albolote (Granada): Editorial Comares, 3-31.
- García González, M. (2013) Free and Open Source Software. In *Translator Education. The MINTRAD Project*. In *The International Journal for Translation & Interpreting Research*. 5-2., 125-148.
- Jordan, P. (1998) *An Introduction to Usability*. London: Taylon & Francis.
- Prior, M (2010) The open-source model. In *ITI Bulletin*, January-February.

