

Technology and translation

(a pedagogical overview)

JOSÉ RAMÓN BIAU GIL & ANTHONY PYM

*Intercultural Studies Group
Universitat Rovira i Virgili
Tarragona, Spain*

*Abstract. Recent decades have seen the work of translators shift into several new dimensions, mainly due to technological advances and the process of globalization. The dramatic increase in the information to be translated, along with the availability of translation-memory tools, has led to changes both in the translator's work processes and in relations with clients. This text presents an overview of these developments, looking at the principles of translation memories, the non-linearity of the information objects translators work on, the corresponding concept of "content", the rise of content management, the use of localization tools, and the role of machine translation. While it is agreed that translation technologies may increase consistency and allow translators to focus their best efforts where they are most needed, the many possible disadvantages include high costs in terms of financial outlay and learning curves, the deepening of divisions within the labor market, and the conceptual restriction of translation to narrow text-replacement activities. It is concluded that the solution to these problems lies in developing greater control over technology.**

Technology extends human capacities. The monkey uses a stick to get a banana, and that stick is technology, in this case a simple tool. More general technologies are collections of tools. Some of them affect our communications, and thus translation.

The use of books rather than scrolls, for example, made it easier to retrieve and cross-reference texts. Concordances were written for complex texts like the Bible, and translations thus had to render the whole text, not just isolated phrases so that the references would work. Similarly, the move from parchment to paper, which was generally cheaper and more transportable, meant that more written copies were made, revised and distributed.

* This paper was written within the frame of the research project "Evaluación de medios de aprendizaje a distancia en la formación avanzada de traductores" (BFF-2002-03050), Ministerio de Ciencia y Tecnología, Madrid.

And since written culture was more easily re-written, translations were commonly re-translated. Not by chance, the use of paper coincided with the translation schools in Baghdad in the ninth century and Toledo in the thirteenth. Or again, the use of print technology from the fifteenth century supported the ideal of the definitive text, hence the definitive translation, and thus notions of equivalence as a relation between stable, fixed texts.

What might we say now that our key technologies are electronic? Texts on the web are constantly being updated, as is our software. We are sometimes called on to render no more than the updates or adaptations. Our translations might thus be expected to move away from the ideal of equivalence between fixed texts, becoming more like one set of revisions among many. In the fields of electronic technologies, translators are less commonly employed to translate whole texts, as one did for the books with concordances. Translation, like general text production, becomes more like work with databases, glossaries, and a set of electronic tools, rather than on complete definitive source texts.

Here we shall be looking at a series of electronic tools that extend human capacities in certain ways. These tools fundamentally affect 1) communication (the ways translators communicate with clients, authors, and other translators), 2) memory (how much information we can retrieve, and how fast), and 3) texts (how texts now become temporary arrangements of content). Of all the tools, the ones specifically designed to assist translators are undoubtedly those concerning memory. But we shall see that electronic technologies affect all aspects of the translator's work.

Translator-client communications

In our digital age, electronic formats concern not just our texts, but also our communications with clients and other translators. Thanks to the Internet, professionals from all over the world can be in regular contact by email or various forms of instant messaging. Work can be sent and received electronically, across national and cultural borders. This has several consequences.

First, in theory, you can work for clients anywhere in the world. The market for translations need not be your city or your country. A source text received at 5 pm in Tarragona can be sent to a translator in New Zealand, who will return the translation before 9 am the following morning, Tarragona time. Time zones can thus be used creatively, and work can thus come from companies that are very far away. All you have to do is list your name, language combinations and areas of specialization on one of the many web sites that aim to put translators and clients in touch with each other. One would expect this process to lead to a situation where the fees paid for translations will become virtually the same all over the world, in keeping with theories of a global market. This, however, is very far from happening.

Translation is still a service that depends on a high degree of trust between the translator and the client. Little constant high-paid work will come from unseen clients; the fees paid in different countries still vary widely; the best contacts are probably still the ones made face-to-face and by word of mouth.

A second consequence of electronic communications is the increased security risk. Translators quite often work on material that is not in the public domain, and this is indeed one of the reasons why relations of trust are so important. When sending and receiving files, you will have to learn various forms of zipping, secure FTP, or other company-specific forms of encoding, with all their corresponding passwords.

A third consequence is that electronic communications make it relatively easy to distribute very large translation jobs between various intermediaries. The client may want to market their product in 15 European languages. They hire a marketing company, which hires a language-service provider, which hires a series of brokers for each language, who give the work to a series of translation companies, who pass the texts on to translators, often freelancers. In this kind of system, the client may be paying as much as four times what the actual translators are receiving per translated page. But each link in the chain is revising, coordinating and producing the various translation products, adding value as they go. This means the text the translator produces is commonly not the same text as the one actually used, and there can thus be little question of copyright over the translator's work. It also means that translators are sometimes very far removed from the end client and the overall context of the texts they work on. Translators in projects like software localization quite often see no more than lists of phrases, along with glossaries that are to be respected. The resulting work can be quite isolating and dehumanizing.

Electronic communications have also been used to enhance communication between translators, especially through Internet forums for professional translators. These are usually classified by topics and/or language pairs. Some may be open, in others participation is restricted to registered members. The traffic (number of emails) in each group varies from a few emails a month to hundreds a day. In these forums translators are very willing to exchange advice, give tips, and generally discuss their work. Simply by reading the posted messages, students and novice translators can learn about translation and see the kind of support that professionals give each other. Discussion lists for professionals usually have their own communication guidelines, and so new participants see a specific way of interacting among professionals. For example, when asking about terminology, professional translators usually send a short message in which they give the term, some context, suggested translations and the consulted sources. This model gives valuable hints about terminology mining and teamwork skills. Or again, by reading messages about a specific computer tool, novice translators often discover that the program is in constant

evolution and has functions they would have otherwise overlooked. These forums thus build a valuable bridge between students and the professional world. They also put paid to the stereotype of the professional translator somehow isolated behind a wall of dusty dictionaries.

Translation memories

Translation memories (TMs) are programs that create databases of source-text and target-text segments in such a way that the paired segments can be re-used. These tools are invaluable aids for the translation of any text that has a high degree of repeated terms and phrases, as is the case with user manuals, computer products and versions of the same document (website updates). In some sectors, the use of translation memories tools has speeded up the translation process and cheapened costs, and this has led to greater demands for translation services. The memories do not put translators out of work; they ideally do the boring routine parts of translation for us.

Translation memory tools re-use previous translations by dividing the source text (made up of one or several files in electronic format) into segments, which translators translate one-by-one in the traditional way. These segments (usually sentences or even phrases) are then sent to a built-in database. When there is a new source segment equal or similar to one already translated, the memory retrieves the previous translation from the database.

An example of the Trados Workbench translation memory suite can be seen in Figure 1. Here we are translating the segment “Restart your notebook” (highlighted in gray); the memory has proposed “Apague su ordenador portátil” as a translation, based on the translation of a previous segment (in fact the one translated just three segments earlier). But “apague” means “turn off”, and here we need “restart”. This is where translators either type a new target sentence or modify the result from the memory database. In this case, we would accept the suggested phrase but change “apague” to “reinicie” (restart). We do not have to rewrite the rest of the phrase.

At the top of the screenshot we see that Trados Workbench has highlighted the differences between each segment and reminds us about the language combination with a flag system. With Trados, we can translate Word documents using the Word itself, but files with other formats need to be translated using specific built-in translation environments.

The platform used by most other translation memory suites (DéjàVu, SDLX, Star Transit) is quite different. Figure 2 shows the user interface of DéjàVu X. Here we have the source text in the left column and the translation in the right one. The suggestions made by the translation memory are in the bottom right corner of the screen. In this system we do not see the document layout, since all the formatting is represented by the bracketed

numbers. Formatting is thus protected. This means that translators cannot alter it by mistake. It also means they cannot edit it consciously.

Translation memories change the way translators work. If you are provided with a memory database, you are usually expected to follow the terminology and phraseology of the segment pairs included in that database, rather than write the text using your own terminological decisions and style. Further, translation memories enable several translators and revisers to participate in the production of the same translation. While this is needed to meet industry deadlines, it may lead to a translation with no cohesive style, made up of a set of sentences put together. The result can read like a “sentence salad” (cf. Bédard 2000).

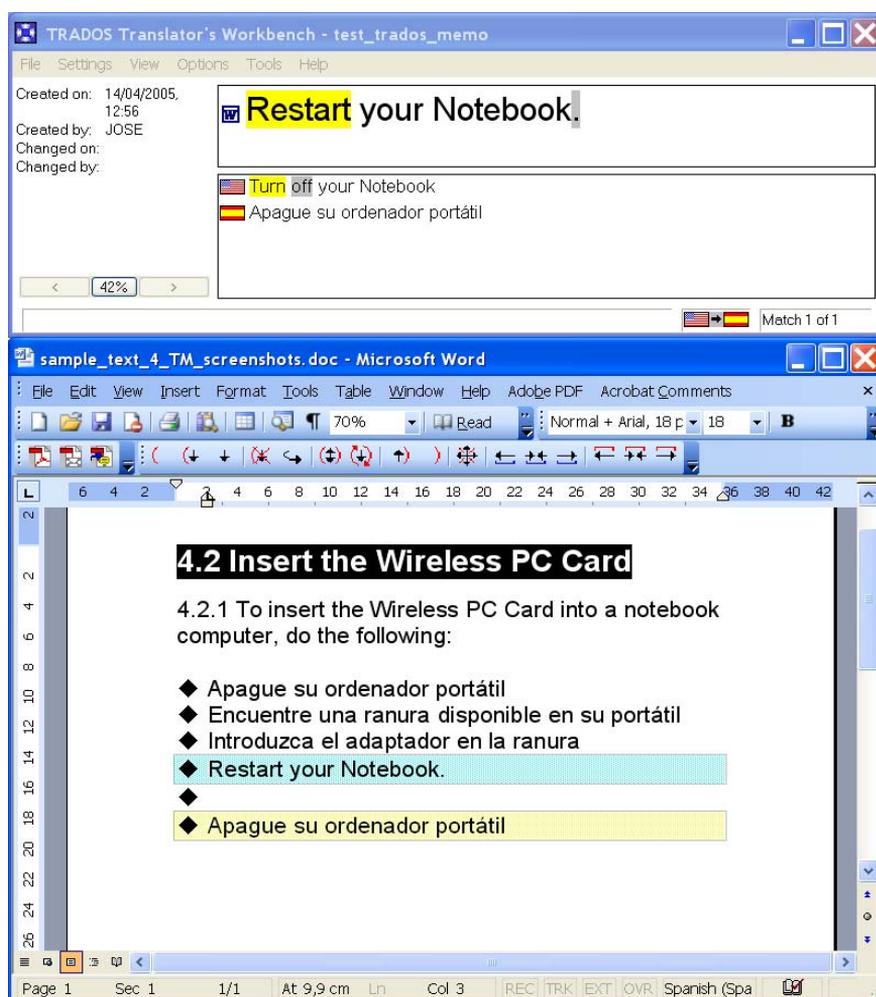


Figure 1. Screenshot of Trados and MS Word

The possibility of re-using previous translations means that clients ask translators to work with TM systems and then reduce the translator's fees. The more exact and fuzzy matches there are (equal and similar segments already translated and included in the database), the less they pay. This encourages translators to work fast and often uncritically with the previously translated segments, with a corresponding decline in quality. When higher-quality work is required, special emphasis must be put on revising the outputs of translation-memory tools.

An associated complication of translation memory software is the ownership of the databases. If you sell your translation, should you also sell the database of matching segments that you have created while doing the translation? Should you sell that for an added fee? Then again, if you have used the work of previous translators by importing a database (or receiving one from your client), can we say that the translation is really all yours to sell? These are ethical questions that escape the parameters of traditional copyright agreements. The possible legal frameworks vary from country to country (cf. Megale 2004). In practice, however, translators receive and deliver databases without paying or charging fees, thus according effective ownership to the clients or language-service providers they work with. At the same time, most translators are used to keeping copies of the databases,

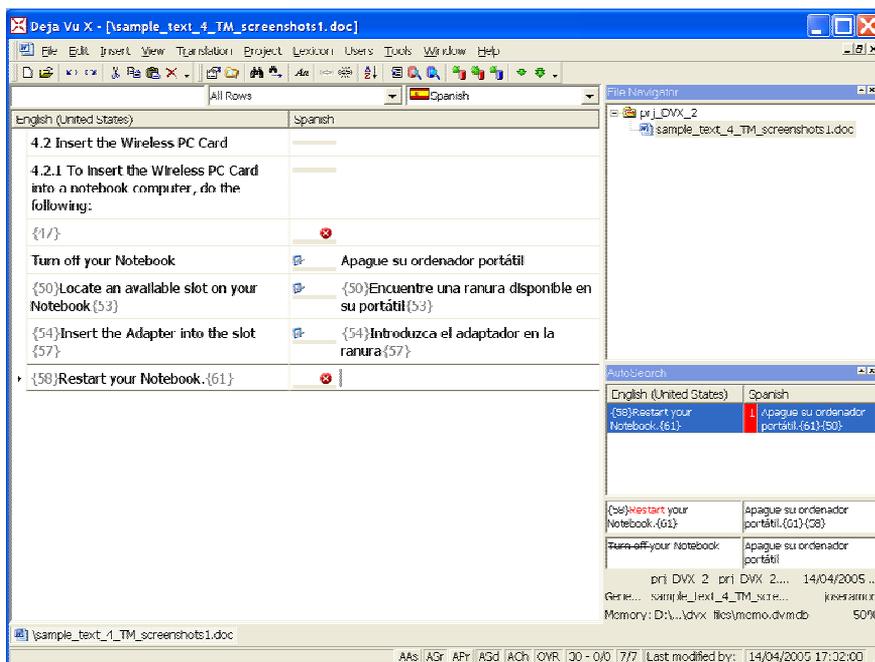


Figure 2. Screenshot of DéjàVu X

or integrating them into their own. To our knowledge, no law has yet been used against them.

This practice, though, will almost certainly die out with the use of the online memories. This system is highly appreciated by clients, since their texts and memory databases remain on a secure server rather than being copied and scattered to translators' personal computers all over the world. Moreover, the owner of the database server (the client or language vendor, never the freelance translator) is the only owner of the memory, as there are no other copies. This means that when these technologies become widespread, translators will not have access to their own previous translations, and project managers will be the only masters of the reference materials translators have access to.

The industrial applications of translation memory tools are based on the idea that translation is a word-replacement activity. On the other hand, translation theories since the 1980s have tended to see translators as communicators whose duties go beyond the replacement of source-text words; translators are employed to provide meaningful communication. Translation memories make this difficult. Indeed, they move translators back to the linguistic equivalence paradigms of the 1960s. Worse, now that texts commonly comprise not only written words but also images, videos and layout (think of any website), translation requires a division of labor. Thanks to our tools, translators are commonly only expected to deal with written words; they are invited to forget about the other elements configuring the text. This division of labor may not always create satisfying long-term employment.

Texts without ends

The way translators work is also being affected by the nature of the texts we work on. We are all familiar with texts that have automated cross-references (links) to other documents, which enable the reader to jump from one text to another. The most common examples are the links in websites. The use of these links means that there is now no clear beginning or end to texts, and that readings are no longer expected to be linear. Indeed, we now talk about "users" rather than "readers". While this is nothing fundamentally new (concordancing is an ancient activity), digital support has radically extended the role of this kind of text.

A major extension can be seen in content management systems. These are computer programs designed to manage databases comprising "information chunks" (generically known as "content"), usually no longer than a couple of paragraphs, which are combined and updated to create several customized texts according to the user's needs. The information chunks are regularly updated and re-labeled. This means that there is no final text, but a constant flow of updated, rearranged, re-sized and user-adapted

provisional texts based on a large database of content in constant change. Think, for example, of a company that produces a series of improved versions of their products, be it software programs or cars, or adapts the products to a series of particular markets. They are not going to rewrite all their user manuals from scratch with each new version. They will logically re-use their existing texts, putting them together and modifying them on each occasion.

XML (eXtensible Markup Language) is a technology standard used to exchange content. It is a way of tagging information so that it can be retrieved later. Take the following example of an XML text:

```
<i tem>
<ti tle>Pride and Prejudice</ti tle> was writ ten by <author>Jane
Austen</author> in <year>1813</year>.
</i tem>
```

```
<i tem>
<ti tle>Alice in Wonderland</ti tle> was written by
<author>Lew is Carroll</author> in <year>1866</year>.
</i tem>
```

By tagging texts as we see above, we can later retrieve information that talks only about authors, for instance, to create a coursebook on literature (in which case we would get both information items). We can also retrieve information based on dates, to create a chronology of publications between 1800 and 1850 (in which case the second item would not appear). With the use of XML in this way, the text production process is anything but linear.

Translating this kind of information cannot be linear either. The updated texts are not translated from scratch, but pre-translated with a translation-memory tool. The translator's duty is to translate only the segments that have been modified, since the non-modified sentences have been retrieved from the memory database. On other occasions, the translator may receive a series of small chunks to translate, usually in a non-problematic format like RTF. These will look like phrases and paragraphs that have no connection with anything. They all have their number or code; they must all respect the established glossaries; they give the translator no indication of how they should fit together. In such cases, translators are obliged to "fly blind", rendering phrases without having any idea of the communicative context.

The development of these work practices has changed the very words used to describe what translators do. Once upon a time, translators worked on source texts, perhaps with the aid of a dictionary. Then, when the importance of contexts and clients was recognized, we talked about "translation projects", involving a lot of background information about specific communicative situations, including specialized glossaries and detailed instructions. In the days of content management, however, it is more exact to refer to "translation programs", on the model of the "maintenance programs" that keep things working day after day, year after year. In the

world of content management, translators may be employed on programs that have cycles, rather than on texts that have ends.

Localization, its terms and its tools

These changes have also brought about a series of new terms for the language industry itself. Most prominently, from the 1980s the need to translate and adapt software to new markets led to common use of the term “localization” rather than “translation”. This term has been defined by LISA (the Localization Industry Standards Association) as follows:

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. (cit. Esselink 2000: 3)

The word “localization” is associated with “locale”, a term to define a specific target market. Locales are often smaller than countries or languages. Localizing a word processor developed in the United States so that it can be sold in the Spanish market involves translating into Spanish the menus, the dialogue boxes and other user-visible messages, translating the online Help files, the publicity and the printed reference material, and adapting any cultural references along the way. But it also involves implementing the word processor with a spellchecker for the variety of Spanish used in the target locale, adapting the “insert date” option so that the text inserted appears as Day/Month/Year, and not Month/Day/Year, including pre-set page settings that match Spanish standards for paper and envelopes, and changing functions so that letter combinations make sense to a Spanish user (the hotkey combination ‘Alt + E’ opens the Edit menu in the English version of Microsoft Word, but the same menu opens with ‘Alt + M’ in the Spanish version, referring to the Spanish word *Modificar*). All that can be called “localization”. It involves more than just translation.

The complexities of localization can be reduced by foreseeing the difficulties and preparing for them in the first version of the product. When this is done, companies save time and money, and may offer better-quality products. This process is called “internationalization”:

Internationalization is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design. (LISA definition, cit. Esselink 2000: 2)

Internationalizing a computer product means designing to handle demands such as the accented characters that will be needed in the localized versions. For example, by designing “Cancel” buttons that are actually much longer than the English word “Cancel”, they allow for longer translations in other

languages (*Annular* in Spanish), so that there is no need to resize the button to display the Spanish translation.

The term “globalization” is sometimes used instead of “internationalization”, notably by Microsoft. For LISA, however, “globalization” involves a specific reference to the way companies are organized:

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 internationalization and product design, as well as marketing, sales, and support in the world market. (LISA definition, cit. Esselink 2000: 4)

We might thus say that globalization is a mode of organization that uses internationalization in order to prepare for localization.

So, is translation part of localization, or vice versa?

The answer really depends on whom you ask. Software developers argue that translation is only one of the many modifications a program has to go through in order to be localized. Translation scholars, on the other hand, might argue that localization is only a fancy name for the act of adapting a text for a specific target readership, which is something translators have been doing for millennia.

Another answer might be found in the electronic tools that have been developed especially for localization. Apart from text editors, spellcheckers, translation memories and terminology management systems, which are common in translation programs, professional localization tools include functions to resize dialogue boxes, assign and check hotkeys to menus, edit and check programming code, manage non-textual resources (such as icons or sound files), calculate the complexity of a project, and replace programming code to make the program work on another platform. The result might still be a translation, but the work process clearly goes beyond traditional translating.

Software localization requires those specific tools. Without them, we would have something like Figure 3, which shows the programming code for a dialogue box. Translating in this format is extremely dangerous and time-consuming, since you can easily delete or modify code (instructions for the computer) by mistake. Before working in this way, you would need to learn to discriminate natural-language strings from code. Moreover, there are no spellcheckers or advanced text-editing tools available.

Are translation-memory tools any better? Figure 4 shows a screenshot of the Transit translation memory suite. Here we have the same file being translated, with source and target text, dictionary and memory database. Translators working with this environment cannot see or edit the code, so their duty is to translate text. Even if they have the ampersand symbol, which stands in front of the hotkey letter, it is the localization engineer who should

test the allocation of hotkeys and perform all the tasks related to layout and function.

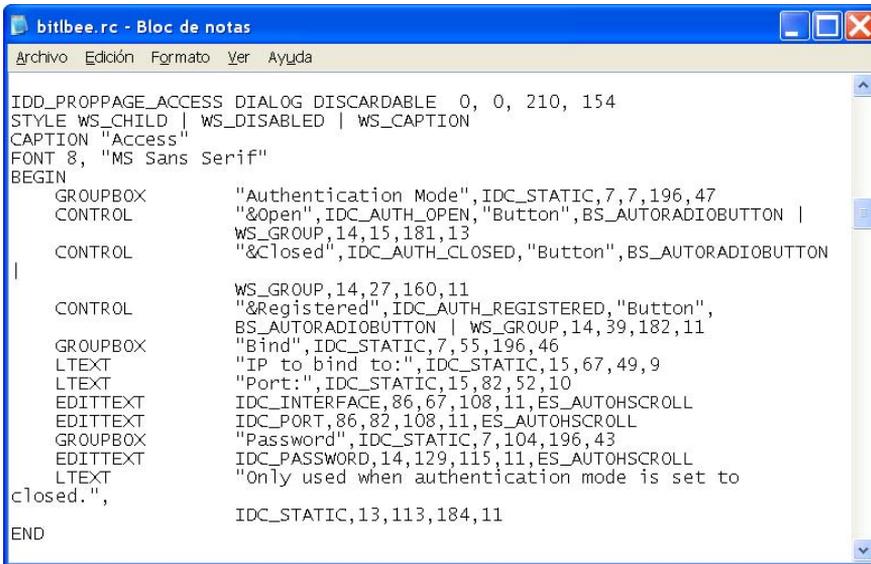


Figure 3. Translation of a resource file with NotePad

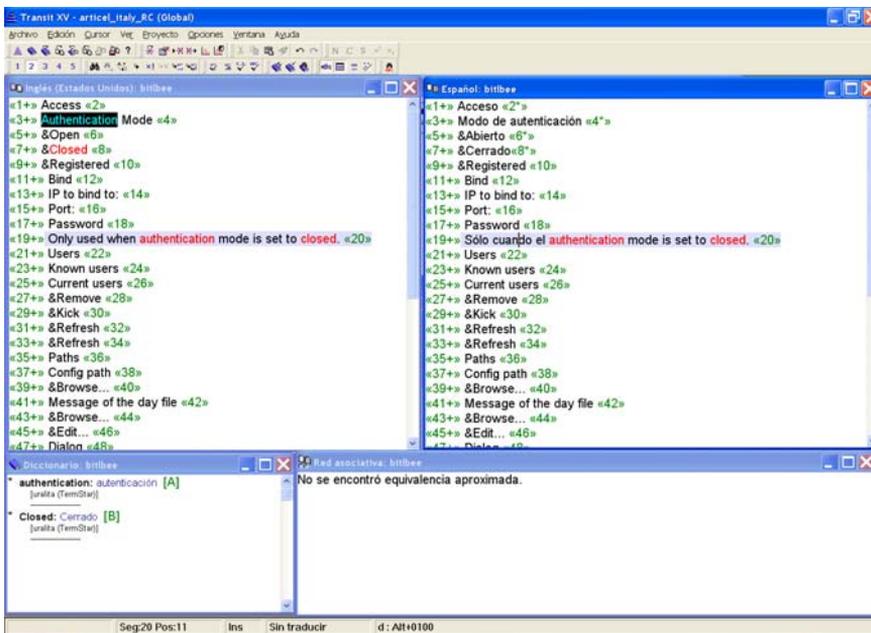


Figure 4. Translation of a resource file using Transit

So what are the tools developed for localizing software? In Figure 5 we can see the same file being translated with Catalyst, which is specifically designed for this kind of work. This is perhaps like the translation memory, except that here we can see what the dialogue box looks like; we have its visual context. And just below the dialogue box we have the active segment (“Authentication Mode”) where we can type our translations and see immediately if the target text fits into the box. We can assign unique hotkey combinations to each function and resize the dialogues if we need to allocate more space for our translation.

Tasks become considerably easier when you get the right tool for the job.

Machine Translation

Machine translation (MT) is probably the translation technology with the most sway over the popular imagination. The first serious attempts to create MT systems date from the late 1940s, when United States and the USSR both funded projects to move rocket technology out of German, and then to spy on each other. It is often said that the initial expectations were very naïve, which would be why when the early projects were almost completely abandoned in the US following the negative ALPAC report in 1966. However, the early approaches were based on quite sophisticated concepts of code-breaking, and there is little evidence that the aim was to produce high-quality output that would be of immediate use. Indeed, the main limitations of the day were on the capacity to store and retrieve huge amounts of lexical, morphological, syntactic and semantic information. The funding evaporated when the Cold War went through a relative thaw.

Several generations later, MT is readily available and relatively functional. The transfer-based Systran system can be used for free on several websites. Its many unhappy matches and almost complete inability to handle proper nouns can result in hours of fun for bored linguists or enthusiastic revision classes. However, the system is extremely useful for gist translations from languages you know nothing about. It allows users to identify the texts or fragments of interest, which they can then have translated by other means.

In other circumstances, MT systems produce high quality translations in very restricted contexts. This can be done by limiting the lexical and grammatical structures of the source text (controlled language) and fine-tuning the system to work only with a specific text type. A classic case is French-English weather reports in Canada, for which an MT system has



been in continuous use since 1984. In other circumstances, a company may develop a highly standardized central language and fixed document templates, enabling MT to be used successfully in conjunction with controlled writing of content (in fact a form of internationalization) and careful revising of MT output (cf. Lockwood 2000, on the heavy machinery producer Caterpillar, where content is written in “Caterpillar English”). The Translation Service of the European Commission similarly uses its own version of Systran to give acceptable results on formulaic texts between cognate languages (especially from French to Italian or Spanish).

There are important technical differences between these examples. The Canadian weather reports and the use of EU Systran are based on correspondences between language pairs (a “transfer” architecture), whereas the use of controlled writing (as at Caterpillar) enables MT to go from one language to many languages at the same time (thanks to an “interlingua” architecture). From the translator’s perspective, however, the consequences are the same.

Machine translation systems are not replacing human mediators. This is first because the prime use of MT is only to *locate* the texts and fragments requiring human translation. Second, if MT output is to be used professionally, it requires human revision. And third, the future development of quality MT output requires serious attention to controlling writing of the input, which is an area that some translators may want to move into. Indeed, the better MT systems work (and current statistical models seem able to offer a better future), the more texts will be processed, and the more work will be created for human translators.

Whatever happens, do not let a client tell you that you have been replaced by a machine. If they say that a text has already been translated automatically and you only have to correct the errors, look for another client. It will usually take you less time to translate from scratch rather than identify and correct the errors. And your quality will be higher.

Advantages and Disadvantages for Translators

Technology is not an option in today’s professional world; it is a necessity. Years ago one talked about Computer-Aided Translation (CAT). That now seems a redundancy. Virtually all translating is aided by computers. Further, the most revolutionary tools are quite probably the everyday ones that are not specific to translation: Internet search engines, spell checkers, search and replace functions, and revision tools have had a huge impact on all forms of written communication. On countless levels, the advantages presented by technology are so great that they cannot be refused. Translation memories perform the most repetitive tasks so that translators can concentrate on the most creative aspects of translation. The intelligent use of machine translation should mean that our best human efforts are focused where they

are most needed. However, technology is not perfect, and translators must be very aware of those imperfections. Here, in closing, we offer a list of those aspects where critical awareness seems most needed.

Each new technology requires new investment, not just in purchasing tools but also in learning how to use them. In all cases, the investment you put in should be less than the benefits you expect to gain. This means, for example, that the kind of text corpora that linguists use in order to study language are generally not cost-beneficial tools when applied to professional translation. They address problems that are more easily solved with a quick web search, and the kinds of quantitative data bases they use have little to do with those developed by translation memory tools. Or again, there is little need to take a course in a particular translation-memory suite if you already know how to use a rival brand. All the products are similar in their underlying technology, and you should be able to find your own way from one to the other. As a general rule, inform yourself before buying anything or signing up for courses. Demonstration versions of all tools are usually available on the web for free, many of them with online tutorials, and translators' forums can give you numerous pointers about the relative advantages and drawbacks of each tool.

Investment in a certain technology can be essential if you are to move from one segment of the translation market to another. The jump is usually made when a client or intermediary offers you work requiring knowledge of a certain tool. You then have to learn very fast, but you are at least sure that you have the right tool for the available job.

Even within large projects, the cost of technology tends to form a set of internal barriers. For example, in a localization project, the project managers (responsible for the overall organization) usually have very powerful, expensive tools with advanced project-management options. The language project managers (responsible for a specific version of the product) have tools that allow them to perform advanced processing, such as automatic terminology checking and the preparation of the packages they send to translators. At the end of the line, the translators have the cheapest and most restricted versions of the software, sometimes called "light" versions, with which they can only replace language strings and perform some basic editing and proofreading tasks. Since code is protected, only the people with the original source files and the powerful tools are able to edit the layout or the content of the source text. By limiting the functions of the tools assigned to each member of the workflow chain, technology has become one way to control the actions and responsibilities of translators.

All these barriers can, of course, be overcome. Translators can and do move into high-tech sectors; some do become project managers, marketing experts, or owners of companies. In general, the way to advance within the profession usually involves more conceptual control over technology, not less. Too often, the dominant industry workflows impose their own specific

technologies and processes. Only when translators are critically aware of the available tools can they hope to be in control of their work.

References

- Bédard, C. 2000. "Translation memory seeks sentence-oriented translator...". *Traduire* 186. http://www.terminotix.com/eng/info/mem_1.htm. Visited 7 September 2005.
- Esselink, Bert. 2000. *A Practical Guide to Localization*. Amsterdam and Philadelphia: John Benjamins.
- Lockwood, Rose. 2000 "Machine Translation and Controlled Authoring at Caterpillar". In *Translating into Success. Cutting-edge strategies for going multilingual in a global age*. Robert C. Sprung (ed.), Amsterdam/Philadelphia: John Benjamins. 187-202.
- Megale, Fabrizio. 2004. *Diritto d'autore del traduttore*. Napoli: Editoriale Scientifica.