Using Information Technology to Optimise Translation Processes at PricewaterhouseCoopers Madrid

Ross Smith

Manager
Translation Service
PricewaterhouseCoopers
Paseo de la Castellana 108
28046 Madrid
Spain
ross.smith@es.pwcglobal.com

Abstract

This paper describes how information technology is used by the Translation Department of PricewaterhouseCoopers in Madrid to optimise translation processes. It commences by describing a mechanism for handling workflow via the corporate network, designed to maximise speed and efficiency in translation requests and also to function as an automated record for administration purposes. This is followed by an appraisal of the CAT system used in the Translation Department, namely the Trados Workbench and related applications. Finally, an ongoing project for making MT (Systran) available to PwC employees around the world over the Firm's intranet is outlined.

Overview of Work Flow

The translation department at the Madrid office of PwC handles translation for all the Spanish offices (20 in total) except Barcelona, which has its own translation team, and for all the Spanish Firm's divisions (Auditing/Business Assurance, Tax and Legal Services, Financial Advisory Services, Management Consultancy Services, Global Human Resource Management and Business Process Outsourcing).

The automated translation work flow commences with the source language document, usually written in Spanish to be translated into English, though sometimes other languages combinations are involved (English-Spanish, French-Spanish-French, German-Spanish-German being the most common). Around 75% of these documents are in Microsoft Word; 15% are in Microsoft PowerPoint and the remaining 10% are in another format or submitted on paper. Most of the paper documents are "sworn translations".

The internal source documents are prepared by secretaries, typists or, increasingly, the auditors, lawyers and consultants themselves. External materials such as articles from newspapers and specialised journals, as well as documents supplied by the Firm's clients, are also translated.

PowerPoint used to be employed primarily for presentations of the Firm's services, in line with its intended function, but it is increasing used for ordinary reports that will be printed and delivered to clients on paper. This slows down translation, though it has little effect on the use of the CAT tools since the PowerPoint documents are rarely appropriate for the Trados Workbench.

Paper documents are sent to the Translation Department by internal courier and evidently are translated in the usual manner, by a "human" translator. Occasionally paper documents are scanned, though this is rarely worthwhile.

On-line Translation Request Form

Documents in electronic format are sent using a Translation Request (*Solicitud de Traducción*) via the Lotus Notes corporate network. This consists of filling in an on-screen form with all the information needed by the Translation Department: name of requester, telephone extension, target language, deadline date, internal charge code, person to whom translation should be sent by Lotus Notes when completed and any other relevant information.

Once the form has been filled in and the electronic document attached if appropriate, the user presses a button marked SEND and the Translation Request is thus submitted. In fact this mechanism is a data base and the request is not actually "sent", but that is the impression gained by the user.

All the Translation Requests are visible to the members of the Translation Department only. Users can track their requests, but cannot view or access other requests.

The manager in charge of the Translation Department assigns the jobs as they come in, on the basis of the language involved, deadlines, translator availability, etc. The translator to whom the request is assigned receives a notification via Lotus Notes and can also see the request under his/her own name in the Translation Requests display.

Use of the *Solicitudes de Traducción* data base to send documents to the Translation Department is absolutely mandatory for the whole of PwC Spain. It constitutes the departmental record of work pending, work-in-progress and completed translations.

So far, therefore, the entire process is automated. For very short jobs, say a few lines or a paragraph, the translator may even decide not to print the text and simply do it on screen, avoiding any use of paper at all.

Once a translation is finished the electronic file, usually in Word, is sent by Lotus Notes to the person indicated in the Translation Request. In the case of sworn translations, the printed documents are sent by internal courier.

CAT Tools

At PwC Madrid the Trados suite is used: Translator's Workbench, WinAlign, T-Window (for PowerPoint) and MultiTerm. References below to "The Workbench" may in fact cover some or all of these tools, not just the Workbench *per se*.

Three different translation memory systems were analysed in depth before a decision was taken to acquire one: Trados' market leader, the Translator's Workbench; Déjà Vu, produced by the Spanish company Atril and popular among freelance translators; and the Xerox Multilingual Memory Suite, which had recently been placed on the market after years of internal development at Xerox, Coopers & Lybrand's main supplier of copying and printing equipment at the time. In the end the Trados system was chosen on the basis of its sound track record and solid technical support service, suitability for multiple users on a network and reasonable price, in addition to already offering an alignment tool and a PowerPoint interface. The decision was not easy, since the other two systems also gave us a very favourable impression.

The Trados Workbench has been in use at PwC Madrid since November 1999. Two members of the Translation Department were assigned to learn to use it (two-day course at PwC by Trados instructors) and try it out over the following year. The results were sufficiently positive to increase the number of users to four in September 2000.

The Department already had two, relatively unsophisticated, in-house translation aids: a large terminology data base in Lotus Approach, with around 5,400 entries, and a search-and-replace macro in AmiPro, the word-processing program used until it was replaced by Word in 1998. The search and replace macro was designed for companies' annual accounts, which contain a high volume of repetition, and was useful despite its evident limitations (e.g. no fuzzy recognition). It became redundant with the introduction of the Workbench.

The terminology bank in Approach was migrated to MultiTerm via Microsoft Access. This was laborious, but the result was a well nourished MultiTerm, kicking off with over 5,000 terms. The number of entries now approaches 8,500, consisting mostly of very specialised terms from areas such as banking and insurance which cannot be found in available dictionaries. We also took advantage of this conversion to cut out a lot of deadwood in the form of duplicated entries, redundant entries, etc.

The two translators working with Trados started by aligning some basic Spanish documents with their existing English translations using WinAlign. These consisted of PwC's sets of standard annual accounts in English and Spanish (individual, consolidated and abridged accounts), comprising the auditor's opinion, the balance sheet, the profit and loss account and the notes to the accounts. The notes constitute the most voluminous part of a set of accounts. A few other documents were aligned then or have been since, including certain standard cover letters, clauses from contracts and powers of attorney, and professional service proposals.

Initial Expectations

The principal documents for which we hoped that the Workbench would be helpful were annual accounts, contracts and powers of attorney. These documents had high levels of repetition, particularly annual accounts with respect to both the accounts of different companies and the same company's accounts from one year to the next. Certain paragraphs, in particular those concerning the accounting policies applied, treatment of taxes, mandatory replenishment of reserves, etc., are identical from client to client and from year to year. This is because most companies use standard headings and paragraphs specified in the *Plan General de Contabilidad* (General Accounting Plan) and in guidelines issued by the Spanish Auditors' Institute (ICAC), which can account for up to 80% of the document in extreme cases. For this reason a search-and-replace macro had been feasible even without a fuzzy component, though it was not fully effective.

In contracts (clauses concerning holiday allowance, severability, notifications, duration, governing law, etc.) and powers of attorney the amount of full repetition is lower than in annual accounts and the use of the Workbench with these documents has been less successful. Simply, they are often very similar, but not similar enough. We continue translating such documents using the Workbench in the hope that the translation memory concerned will eventually become large enough to be productive.

Annual accounts are very Workbench-friendly, as already mentioned, in three basic ways: large similarly for same company from one year to the next, large similarity in accounts of different companies, large similarity in individual and group (consolidated) accounts. For instance, when translating the individual accounts of a parent company, say Acerinox, and also the consolidated accounts in which the financial statements of all the Acerinox group companies are brought together, once the individual accounts are done and loaded in the Workbench memory, only about 15% of the consolidated accounts will actually need to be translated: the rest is identical to the individual accounts and is therefore translated automatically. And the following year when the new accounts come in, an even higher level of 100% matching will be obtained.

Implementation

The translators assigned to Trados initially translated all accounting documents using the same main memory, but this rapidly became cumbersome and caused problems of speed and RAM saturation. Therefore a new procedure was implemented, whereby each set of annual accounts was given its own translation memory bearing the name of the PwC client concerned (e.g. Real Madrid). When the document is received, it is first pre-translated using the main memory to translate as many standard terms and sentences as possible. This can be done by applying only 100% matches, in which case the translator can be almost certain that the matches found are totally reliable and concentrate on what is new, or by applying a lower

percentage (say 80%), in which case more text is translated but it needs to be edited with greater care. This is a question of personal preference. Then, a new memory is made and the document is translated sentence by sentence, the already-translated sentences being uploaded to the memory directly or after a quick review, the new ones being "manually" translated and then uploaded. This is slower than doing a fuzzy translation of the whole text, but the advantage is that when the annual accounts of the same client are received the following year, the translator simply runs that client's memory on the accounts and, in principle, will translate up to 80% of the document at the press of a button.

The final stage is to print the document and check it on paper, basically because certain problems are more easily detected that way, particularly errors in formatting.

Productivity gains - i.e. time savings - range from 10%-30% in the case of new documents to 50%-75% in the case of "repeated" annual accounts and other near-identical documents (e.g. employment contracts for two or more different employees in which only a few conditions are altered). Evidently, at first the gains were sometimes inexistant as the users became accustomed to the program, which for a beginner is pretty complicated, but in general even at the early stages very little time was actually lost. After a few weeks the number of user errors declined almost to zero and the gains in time became very apparent. It must be borne in mind also that the longer the Workbench is used, the greater the benefits it provides.

An idea of the times involved can be given in the following example.

Assignment: Spanish-English translation of the annual accounts (balance sheet, profit and loss account, notes to the accounts) and auditor's opinion of a company engaging in a normal kind of activity which does not cause special terminology problems.

· *Length of document: 3,500 words.*
· *Competent translator working fast: about 7 hours.*
· *Experienced translator working fast: about 4.5 hours*
· *Experienced translator using Trados Workbench and working fast: about 2.3 hours*
· *Experienced translator using Trados Workbench and working fast when the same client's accounts were translated the previous year: around 1.5 hours.*

It can therefore be seen that in favourable circumstances, an experienced translator using the Workbench can do the same job in one third of the time that an experienced translator would need without this CAT tool.

According to the record kept in the translation department, during the first full year of using the Trados Workbench (1999-2000) the average gain in productivity for eligible translations was around 25%, including early jobs where gains were minimal, with about 300,000 words (equivalent to 1,200 pages) being translated using the Workbench. During the period July 2000 to June 2001 the number of words translated leapt to 750,000, due to the larger number of Trados users (four instead of two), their greater experience and the availability of well-stocked translations memories.

The Trados suite allows the termbank MultiTerm to be used in conjunction with the Workbench. In our particular case we do not use this option much, probably because all the translators are very experienced and know the terminology well. We do use MultiTerm as our sole terminology data base, and find it extremely useful. As already mentioned, the WinAlign tool is also used occasionally, and was used a lot at first.

The T-Window for PowerPoint is used from time to time, particularly when a PowerPoint documents shows a large amount of repetition. It is also used to make the process of translating such documents simpler. PowerPoint presentations often contain text in little boxes and bubbles all over the screen, sometimes even written vertically or diagonally. This slows down the translator's work and causes frustration. The T-Window helps by extracting all text and showing it clearly in a window, with another window below for the translation. This makes translation a lot easier and therefore the T-Window is sometimes used even when a document contains no repetition. The document analysis function provided by T-Window is also used to obtain reliable word counts for PowerPoint documents, whose own word counts are notoriously unreliable.

In conclusion, the investment in the Trados CAT tools has been most worthwhile. In addition to raising productivity - albeit only in certain areas - it has provided department members with a new option and the chance to learn a useful new skill. Not having to do repetitive, tedious tasks has been viewed with considerable satisfaction, and simply being able to do something different and evidently advantageous has helped to make the translators' work more enjoyable. That is very important.

On-line MT Application for PwC Personnel Worldwide

Briefly looking beyond the Madrid office, a pilot MT system is available to PwC employees worldwide consisting of access to a Systran on-line machine translation application. This initiative was implemented by members of the Firm's Global Knowledge Management team with the aim of making rudimentary translation available to everyone, everywhere. PwC has an entreprise-wide license with Systran that allows staff to access the special PwC/Systran pages, either via PwC's intranet ("KnowledgeCurve") or via the WWW. Users can choose from a number of specialised dictionaries which are designed to raise the quality of the translation. Twenty language pairs are available. Results to date have been mixed, according to user feedback.

As is normally the case with serious MT applications it will be necessary to tailor the system to a high degree for it to be genuinely useful and work in this area is now under way via the inclusion of specific PwC dictionaries in addition to the available Systran ones. Specialist English-Spanish glossaries supplied by the Madrid translation department covering various areas of financial and legal translation are now available on the PwC/Systranet site and it is hoped that PwC's Paris and Amsterdam offices will contribute similar material for French, Dutch and German in the future. Even then its potential will be limited because of the nature of the documents typically used in this kind of organisation, which in general are the very opposite of what MT handles well. Users are warned of this, and are told not to expect it to be a substitute for expert "human" translation. However, it could be of some use for translating, or at least "gisting", emails of a general character and other

uncomplicated documents, and providing some kind of translation service for the more far-flung offices (the office in Mauritius say they use it for French-English translation, for instance).

Outlook at PwC Madrid

Further automation of the translation process at PwC Spain is not likely, there being little left to automate. The implementation of a pure machine translation system is not envisaged at present due to the limitations of such systems and the lack of simplicity and uniformity in the vast majority of the documents translated. The Systran application described above will remain available for anyone who wishes to use it.

We shall continue to use the Workbench as much as possible because of the productivity gains it offers and because it makes the translators' task more varied.