

# **Workflow, Computer Aids and Organisational Issues.**

*Margaret King*

## **Introduction.**

The burden of this article is that since translation services and agencies can vary enormously in the kind of work they do and how they do it, the introduction of electronic documents and the tools that make use of them into the translation process needs to take account of the differences. In particular, the consequent changes in work flow patterns may be very different, ranging from doing little more than offering another possible way of doing things at some point in the translation process to radically changing the way work is divided and tackled.

The ideas put forward here are based on studies carried out for the Translation Services of the European Commission and for the Linguistic Services of the Swiss Federation, as well as on work in progress with the World Intellectual Property Organisation and in the context of a European project, TransRouter. The first three are all concerned with ways of introducing computer aids into existing translation services and with the consequences of doing so, while the last aims at developing a software tool which, on the basis of document characteristics and other constraints, will help the translation manager to decide how best the task of producing a translation should be accomplished - whether, for example, it lends itself to treatment with a translation memory system, whether it can be adequately dealt with by machine translation, whether a particular terminology resource would be helpful, whether it is best dealt with by human translation and so on. The author is however solely responsible for the contents of this article, and nothing said here should be construed as being the official policy of any of the bodies mentioned.

## **Translation Scenarios.**

The article is structured around examination of three different translation scenarios. Once again, while based on real experience, all are over-simplified pictures which, while reflecting more or less faithfully some aspects of their translation work, make no claims to be an accurate representation of any particular organisation.

In each case, a brief description of the scenario is given, followed by some suggestions about where translation technology might be introduced. Finally, in each case, we ask how the introduction of new tools and new ways of working might change the working lives and habits of those involved.

### **Scenario one: a small homogeneous unit.**

*As it is.*

First, consider a fairly small group of about ten translators. A major part of their work is legal translation, systematically adding to the parallel text versions of a growing body of legislation. Although this work must be accomplished within reasonable delays, they are not subject to extreme urgency. Consistency is of very great importance: not only must

new translations be internally consistent in the way they deal with terminology and phraseology, the body of law must be consistent across time.

All members of the group are used to dealing with electronic documents. They work on PCs, have access to a local terminology base which they share with a number of other translation services and also have access to Eurodicautom. They have limited access to dictionaries on CDROM (not all translators have CDROM readers, not all useful dictionaries are available). They do not however use any translators' aids tools other than those associated with a word processor, of which the spelling checker is the most heavily used. Translations are delivered in electronic form. The idea has been mooted that this being so, the group might produce camera ready copy rather than the electronic version being sent on to a publisher.

All members of the group are staff translators. They work in offices on the same floor, conveniently grouped around a central area which contains much of the reference material they need in printed form. It is therefore very easy for them to consult one another, and chance meetings in the central area reinforce ease of contact.

### *As it might be.*

The scenario sketched above suggests the potential utility of several computerised tools. Here we will concentrate on just three.

A preoccupation with consistency in translation suggests that a **translation memory system** could be very useful, especially if the memory contained the body of law which had already been translated.

Such a memory is likely to be large, perhaps very large, and if searching times are not to become unacceptably long, the question arises of how the memory should be structured to avoid this. In the particular context outlined here, the hierarchical nature of the body of law may provide an answer to that question. In other contexts it may not be so straightforward.

Organisational issues also arise: who is to build the memory, how is it to be updated, who will decide what translations are fed into the memory and at what point? Building the memory involves the more or less fastidious task of aligning two texts and checking the results. How time consuming and annoying construction of a translation memory is depends essentially on two factors: how reliable the alignment has to be, and the number of formatting mismatches between the two texts. Two examples from recent experience will help to show how sensitive aligners are to differences of formatting. In a large document, it turned out that a table present in the English version had been omitted from the French version. Not surprisingly, this gravely perturbed the alignment of all the text that followed. But, since the document was long, spotting the disparity between the two texts would not have been all that easy. In the second case, the layout of the title of one version had been done using hard carriage returns, in the other it was one continuous line. This resulted in different segmentations of the two texts, and, again, in false alignments. These kind of problems can become heavily burdensome when a memory is being constructed from existing texts where, naturally enough, page layout and formatting on each version has been done independently by two different people, with the sole consideration of how the text will look on the page in mind.

In some cases, it may be enough to have most of the alignment correct. With our second example above, accepting the alignments given would only have resulted in one segment being missing from the memory, which would not have perturbed overmuch its functioning. In other cases, though, very reliable alignment is needed, and then each alignment must be checked. Although this task does not necessarily require extensive or profound knowledge of the two languages, it is time consuming and not very interesting work. There is an obvious intimate connection between formatting problems and the amount of time needed to check alignment accuracy.

In the context of our first translation scenario, it is likely that very reliable alignment is required, and that the task of creating a memory will be fastidious and time consuming.

Updating the memory may raise additional organisational issues. Most commercial translation memory systems seem to have been designed from the view-point of an individual translator, who can decide for himself what translations to put into the memory when. He may indeed decide to update the memory sentence by sentence as his work proceeds, thus facilitating the task considerably. Once the context is that of a large organisation, mechanisms need to be put into place for validating the translations to be added to the memory and ensuring that the updating is done. We shall return to this issue below in the context of terminology management.

Most translation memory systems come packaged with a **concordancing facility**, which allows the translator to search for all occurrences of a word or phrase in the memory and view them in context along with their translations. This facility is of obvious interest in a context where consistency across translations and across time is important.

With the same need in mind, the group could well benefit from an **electronic archive** independent of the translation memory. The archive would contain not only the body of law but also a substantial amount of related material, and would support a focused search through parallel language versions. For example, the translator could ask for all instances where term is *not* translated in a specific way, or where two or more terms co-occur. This provides an easy access to reference material as well as a tool for checking consistency. And, of course, such an archive would be useful to a considerable wider community of users.

Once again, however, the construction of such an archive raises issues of how it should be structured, created, and maintained.

### ***Consequences on working patterns.***

The use of word processors has already radically changed workflow patterns. There was a time when the translators dictated their translations, which were subsequently typed by secretarial staff. With the introduction of PCs secretarial staff have been much reduced, and the translators type themselves. They do not, however, as yet, have to be overly concerned with formatting of text or of page layout. This will come if they are eventually obliged to produce camera ready copy. At that point, either the translators will have to acquire a new set of skills or secretarial staff will have to be increased again in order to cope with the new kind of work being asked of the service.

It is worth labouring this point a little. Although the chain of producing a draft, having it typed, correcting the draft, having it re-typed, repeating the process if necessary until an acceptable final version is produced, then sending the final version to a publisher where it

is probably re-typed is clearly wasteful of time and human resources in an electronic age, it is only recently that translator training has begun to include the acquisition of word processing skills. Maximum efficiency is unlikely to be gained by requiring people with one set of skills to exercise a different set for part of their time.

Other effects on working patterns depend critically on the answers to some of the organisational issues raised above. The questions connected to the creation and maintenance of an electronic archive will be left aside, on the grounds that a single small group of translators is very unlikely to be charged with this task. For the rest, let us imagine two extreme situations.

First, imagine that the creation of the translation memory is contracted out, and that maintenance of it is entrusted to what we might call a translation technician who, while not being a translator, is very familiar with translation technology. When a document for translation is received, the technician makes sure that it is in a suitable electronic form and extracts from the central translation memory a memory tailored to this translation. (Most translation memory systems offer facilities for doing this as part of the package). The translator receives the text to be translated and the tailored memory, which he can then modify if he wishes during the translation process. The translator does not modify the central memory in any way. What changes in the translator's life is that instead of using a naked word processor, he has access to translation memory and concordancing software through his word processor. He needs to learn how to interact with this software, but that is all. When the validated version of the translation is ready it is passed back to the translation technician who updates the central memory.

A rather less attractive scenario is to imagine that the translators themselves are responsible at least for updating the memory if not for creating it, and that they have direct access to the central memory for this purpose. Two versions of how this might be done are possible.

Most translation memory systems automatically update the memory during the translation process. In the first scenario above, the translator was not working with the central memory, so his new translations were only going into the tailored memory created for this translation. One might imagine that updating could be accomplished by allowing the translator to work directly with the central memory and taking advantage of the automatic updating facility. The translator would then have to learn how to edit the translation memory in order to correct translations automatically entered (every translator occasionally changes his mind, or realises he has made a mistake), and even so the risk of introducing incorrect translations into the memory increases, as does that of damaging its integrity by accident.

Most of those risks can be avoided if the translator works with a copy of the central memory or learns how to create an independent tailored memory for himself. Updating the central memory is then done only after all revisions have been incorporated into the translation and the validated version produced. This may mean that the translators now also have to learn to use an alignment tool and spend some of their time validating the results of alignment. Alternatively, the translator has to 're-translate' the document in order to make use of the automatic updating facility.

## **Scenario two: a slightly larger, more disparate group.**

### *As it is:*

This group consists of about twenty people, most of whom are freelance and part time. They deal exclusively with highly technical documents, each of which is on average about half a page long. Sentence structures are limited, but coordinations and relative clauses abound. Documents contain a great deal of terminology, much of it new terminology. Between 1,200 and 1,300 such documents arrive for translation each week. Deadlines are known well in advance, but the freelance translators are encouraged to keep up a steady rate of production by being paid piece rate.

Translators choose their own working method. The majority type their translations onto PCs, although a few prefer to dictate and one or two hand-write their work. All work arrives on paper. Those translations which are produced electronically are delivered electronically. Dictated or handwritten translations are sent to a separate typing pool for typing. The same typing pool is responsible for copy typing the original documents and for up-loading them into a central data base, where the translations also are stored. Experimentation with acquisition of the originals in electronic form has recently started, but paper can be expected to be a major medium for some time to come. The group has recently acquired access to the term bank TERMIUM, and is very enthusiastic about its helpfulness. They also have access to a specialised data base containing material of the same document type as that which they treat.

Not all translations are systematically revised. When revision is done, it is usually done by the permanent staff on paper copies. The corrections are then transmitted to the translator who incorporates them himself into the final text, or corrections are made in the typing pool.

Although the majority of the staff are freelance, for security reasons they work on the premises of their employer. The unit is rather overcrowded, with the result that a wide central corridor now houses both the library and a desk with work stations. Contact between staff who are present is so easy as to be almost unavoidable, but given that many are part time, not all staff meet all other staff regularly.

### *As it might be.*

The most obvious and pressing need for this group is for easy access to existing terminology resources and for tools to assist with the acquisition and management of in house terminology.

Some projects in this direction are already under way. It has long been the case that paper dictionaries can be acquired easily and quickly, and this is now being extended to dictionaries on CDROM. As mentioned, access to TERMIUM is available, although not all translators as yet have a direct access from their own screens. Access to the Internet is being prepared, with special bookmark files provided that will lead directly to potentially useful sites. A word of caution is in order here, though: although much terminology is available on the net, it is of very uneven quality. The provision of special bookmark files can be used to steer the translator towards good quality resources, as well as avoiding a lot of unnecessary surfing.

A **local terminology management system**, accessible through the translator's word processor springs to mind as the translator's aid most likely to dramatically facilitate terminology research and management. However, introduction of such a tool, as always, raises technical and organisational issues.

First, the advantage of a terminology resource coupled with a word processor is that the translator can search directly from the source text and can incorporate the solutions he finds directly into his translation. If the source is not in electronic form, then the first of these advantages disappears. Thus, an essential first step is to transform the written source documents into electronic form through scanning or perhaps dictation software. An extra task is thereby added to the work flow, that of checking the results of scanning or dictating. It is also perhaps worth noting that even if the originals are available in electronic form, many translators will prefer also to have a paper copy available, partly because it is easier to keep an overall view of the document from a paper copy than from scrolling on the screen, and partly because it is far more comfortable to move the head around, looking sometimes at the screen, sometimes down at the desk than to look fixedly at the screen all the time, getting backache and headache while doing so. Ergonomic considerations are of great importance in ensuring acceptability of any electronic tool.

The second advantage is only a real advantage if the local terminology resource is rich enough to provide a sufficiently large number of successful hits. The translators estimate that TERMIUM currently provides a solution in better than 50% of all cases. If the local terminology source cannot at least approach the same success rate, the translators will soon go back to consulting TERMIUM, copying down the solution or printing the fiche and typing the solution in.

The underlying problem here, of course, is that of terminology acquisition. An obvious source of terminology is the translations themselves, but it is unrealistic to install an empty terminology resource and ask the translators themselves to enter terms as they work - especially when the translator is being paid piece rate. It seems likely then that another essential preliminary to installing a local system is to invest in the creation of resources to stock it.

As with the translation memory system discussed in the last section, updating and maintaining the local terminology source will raise issues of validation and of organisation. Most terminology management systems allow the translator to update the term bank as he works, and provide tools to make it easy to do so. In an organisation, encouraging translators to update the term bank obviously helps to solve the problem of acquisition of new terms, but at the same time runs the risk of damaging the integrity of the term bank itself. Multiple entries may be created, sometimes justifiably (terms do change over time or over subject matter), sometimes not. Where more than one solution exists, the poorer one may be chosen through inadvertence or ignorance and perpetuated through its existence in the term bank. All this suggests the wisdom of putting into place a validation mechanism, and centralising the maintenance and updating of the term bank. But centralisation once again contributes to making acquisition more difficult.

Essentially, two options exist here. The first is to persuade the translators to collaborate and to make it very easy and quick for them to do so, perhaps by asking them to do no more than to mark new terms with fluo pen on a finished original or translation and pass the marked copies on to the central updating section. The other alternative is to provide

those responsible for updating with tools to help identify new terminology from the results of translation.

Although we have mostly concentrated in this section on terminology needs, let us take advantage of the fact that some of the translators in this group still dictate their translations to suggest that their work might be facilitated by the introduction of **voice dictation software**. These softwares have made a great deal of progress in the last few years, and even though training for individual voice patterns is still required in order to obtain the best results, once the training has been done the results can be surprisingly good. But use of dictation software raises another set of issues, this time connected to physical layout of the working place and to ambient noise. The microphones used are quite sensitive, and even with a robust system a ringing telephone can cause some chaos. Space constraints prevent us from going into these issues here, and in any case, office layout and noise conditions are very specific to particular locations. Nonetheless, given that use of dictation softwares promises to have a profound effect on translation work, it is worth being aware that their introduction also brings with it new problems to be solved.

### *Consequences on working patterns.*

As with our previous case, how working patterns are affected depends to a very large extent on answers to some of the management and organisational questions. Once again, let us imagine two alternative situations.

First, let us imagine that every translator now has access to a rich local terminology source through his word processor, as well as to external resources and to the web. These latter can be accessed without having to close the word processor. Care has been taken to ensure that interfaces are easy to use and pleasant to look at, ideally offering the translator a choice of window arrangement, of screen colours and so on. The local terminology source is regularly and rapidly updated centrally with new terminology culled from all the translators' work.

Original documents either arrive in electronic form or are pre-processed into electronic form before the translator receives them. When a translation is finished, an electronic copy of it and of the original is automatically available from a central data base for use in terminology identification or updating of other linguistic resources.

Preparing a translation now becomes mostly a matter of sitting at a screen. There will still be recalcitrant cases where colleagues will be consulted and library searches made, but they will be much less frequent than before. This has some clear advantages: it is quicker to prepare a translation, consistency across translations is enhanced, duplication of work is minimised. But it also has some disadvantages in that social contact between translators is cut down, and the physical effects of spending long periods working on a computer are known to be disagreeable. Awareness of the disadvantages is a first step towards countering them.

The other scenario involves the translator either having to scan originals in himself, or, more probably, receiving the raw results of scanning and having to check and if necessary correct them before he can start work on the translation. He is also expected to feed any new terminology he defines into the local terminology management system, either whilst he is doing the translation or whenever he finishes a translation. The disadvantages here are that the translator is being asked to spend his time on word processing rather than on

translation, and that the integrity of the term bank is in danger. On the other hand, making a direct contribution may mean that the translator becomes more aware of how improving the contents of the term bank facilitates later translation work.

### **Scenario three: one department of a large translation service.**

#### *As it is:*

This is a department responsible for translation into a large number of languages. Internally, it is organised into smaller groups of about a dozen translators responsible for translation into one specific language, but parallel translations are often required into several languages at the same time. Requests for translation come from a large number of different sources. They are roughly linked in subject matter, but come in a very wide range of document types, ranging over a number of important dimensions. Documents can be very brief, a page or less, to very long, several hundred pages. They can be highly confidential, somewhat sensitive or not sensitive at all. They can be information or publicity material destined for a wide public or extremely technical. They can be legal texts, where every translation has legal force. Translation requests can be one-off, or documents may come in a series of versions with each version differing more or less substantially from the previous version. The urgency with which translations are required can range from immediately (within the next couple of hours) to several months down the line.

Most, but not all, documents arrive in electronic form. All translations are delivered in electronic form. Individual translators choose their own way of working: some dictate, with the translation then being typed within the language group, others work directly onto a PC. Translators have access to EURODICAUTOM and to the Internet. The department makes use of freelancers for a substantial amount of translation (over 20%), and typing too is occasionally contracted out by groups which have difficulty in finding or keeping typing staff. The in-house translators all work in the same building. Freelancers work from home or in translation agencies and may be geographically located a considerable distance away. Decisions on when and how to make use of freelancers are typically made within the group responsible for translation into a specific language, although freelancers are centrally recruited and the list of freelancers is centrally maintained.

Substantial support is available to the in-house translators. Help with terminology or with specific language problems is available, as is computer support. Some translation technology tools have already been introduced: in-house translators have access to a machine translation system, as do staff outside the translation service, local terminology management systems have been introduced, there is easy access to a large central terminology bank, translation memory systems are being introduced. Individual translators make use of translation technology tools if they wish to do so, and usage varies greatly.

#### *As it might be :*

The most striking thing about this scenario is how different it is from the two preceding ones. The work to be done is much more varied, the size of the operation is much greater and translation management is correspondingly much more complex. And herein lies the moral of our tale. If we think about how translation technology can be of help to this group, it is immediately obvious that different tools will be of use in different contexts. It



would be impossible to draw up an exhaustive account here, so let us be content with some examples.

**Dictation software** may be useful to someone who has to translate a short but very urgent document. In the same context, a **translation memory system** will probably not be useful, just because too much time is needed to prepare the document and a tailored memory.

When a document comes in several versions, **document comparison software** may be very useful to spot the differences between the new version and the preceding version : even if the requester is supposed to have marked the differences he may have forgotten to do so or not done so completely. In contrast, document comparison software is pointless when only one version of a document is ever going to exist.

This group sometimes has to deal with very long documents where time does not allow the document to be translated by a single translator. In this context, **tools which support group work** are potentially very useful. Some of the same tools might be useful to translators in different language groups when translations of the same document into different languages are being produced in parallel. They are simply irrelevant to someone who needs to work quickly on his own.

Extensive use of freelances chosen from a centrally maintained list raises questions of **quality control** in a particularly acute way. It is no longer possible to get to know the freelances and learn who can be trusted to produce high quality work without revision when it proves to be impossible to revise all work. Could tools be developed that would help in signalling potential poor quality ?

Even on the basis of this very small sample, it becomes clear that different documents need to be treated in different ways, and that it is a combination of document characteristics and external constraints such as deadlines or particular language combinations that help to determine what the appropriate treatment for a document is. It also becomes clear that translation technology is not only about translation aids, it is also about translation management tools.

### ***Consequences on working patterns.***

Consciously deciding that different documents need to be treated in different ways means that workflow itself is document oriented. To illustrate this, let us again take just a couple from the wealth of possible examples.

First, let us imagine that a series of documents (a translation dossier, let us call it) falls into a category where machine translation has been known to give good results, perhaps because effort has been invested in specialising the machine translation system for this type of document, and where each document in the dossier re-uses material from other documents in the dossier. Let us also imagine that a large central translation memory already exists.

A possible way of treating this dossier would be for someone - preferably not the translator, as already discussed above - first to check for possible formatting problems, spell-check the originals as a precaution, and extract a tailored translation memory from

the central memory. The same person could check in a central electronic archive for missing pertinent reference material and extract it from the archive.

The translator would then receive a packet containing documents to be translated, a tailored translation memory and reference material. Choosing the first document to be treated (using as criteria characteristics such as a high degree of internal repetitiveness or that substantial elements of the document are likely to be repeated in subsequent documents), the translator first asks for a machine translation and checks the results. Satisfactory results are used to feed the tailored translation memory, and the rest of the document translated using the memory interactively, that is the translator updates the memory as he moves through the text. This operation is repeated for each document in the dossier. When the translation is complete, it is sent for revision. Revisions are incorporated into the text, and the final text used to update the main translation memory. With his example, it is not strictly necessary that all documents arrive for translation at the same time, it is sufficient to recognise that a document is probably the first of a series.

With this first example, translation is still a black box to the outside world: the individual translator's way of working has changed considerably, but for the requester he still sends a document for translation and gets a translation back, without being aware of how the translation has been produced.

If the document to be dealt with is a very long document, multi-authored with different parts of the document becoming available at different times, one might imagine a workflow that would break open the black box. Here, as soon as it is known that the document will be produced, a team of translators is assigned as a sort of task force to the development and translation of the document. They can be consulted by the authors as work progresses (for terminology or language questions for example - some of the authors may not be writing in their own language) and can start work on producing the translation of different parts of the document as they come available. The author/translator team uses e-mail and groupware to support collaborative work. The translation of different parts of the document is tackled as each translator sees fit.

The main point here, and one which deserves to be laboured much more than space permits, is that workflow does not just concern the translation activity alone, but the whole context in which that activity is done.

## **Conclusion.**

The main conclusion to be drawn from all this is quite simple. Some translation technology tools and some resources are likely to be of benefit to almost everybody. It is hard to imagine, for example, a translator who could not benefit from **a uniform interface through which a wide range of different dictionaries could be consulted**, instead of having to familiarise himself with a new interface every time he buys a new CDROM, or a translator who does not rejoice at having a rich terminology source readily available and easy to consult. Other translation technology tools are best suited to specific situations and to different contexts of work. A critical factor in deciding on the potential utility of the tool is the type or types of documents to be dealt with and the external constraints on translation production. Maximum benefit from introducing translation technology can be gained by careful preliminary analysis of what is really needed and of the consequences of introducing it.

A secondary, but nonetheless very important, conclusion is that introducing translation technology into the workflow introduces also new tasks, which are in many cases not best tackled by translators. In several places in this article, we have imagined the existence of translation technicians as support staff : it may be that a new profession needs to be born.

### **References.**

EAGLES Evaluation Group (1995). Final report. Available from CST, Njalsgade 80, DK 2300 Copenhagen; or electronically at  
<http://www.wissco.unige.ch/projects/ewg96/ewg96.html>

King, M. (1995). The European Commission Translation Service : A Case Study. Available from the author or electronically as an appendix of the report cited above.

