The automated translation of software, particularly the computer user interface and user manuals

Mike Scott

Customer and Service Education, Rank Xerox, Welwyn Garden City, UK

INTRODUCTION

Much has already been said in previous papers about Xerox as a vendor; this paper is about Xerox as a customer of machine translation systems. This is the 10th annual Translating and the Computer conference and this is about the same length of time that Xerox has been involved in using machine translation. Xerox made large investments in the late 1970s but as competition in the marketplace has increased the money available for work in this field has reduced considerably. However, it is worth noting that these large investments have shown results over the last ten years and will probably do so for at least the next five, so the life cycle for these systems appears to be long.

This paper is about large-scale translation in a corporate environment. I do not like quoting statistics as invariably there is always someone to dispute them, but we are probably the largest commercial user of machine translation (MT) or machine-aided translation (MAT) systems in the world. Our annual throughput of machine-translated material is approaching 20 million words or 100,000 pages. Nearly half of this is undertaken in the United Kingdom and in total is still a small part of the total requirement for technical translation in Xerox.

A real product example, the 5090 copier, can be used to illustrate the way the translation process is designed and managed. The year 1988 marked the 50th anniversary of xerography. The xerographic process was invented by Chester Carlson and the first image was produced in Astoria in 1938. The Haloid Corporation turned the invention into a product; Haloid subsequently became Xerox Corporation, which in turn formed

joint ventures outside the United States including one with the Rank Organisation to form Rank Xerox. To commemorate the anniversary, new products launched in 1988 form the '50 series' product range.

The 5090 (see Figure 1) is the flagship of the Xerox fleet: it can copy 135 A4 sheets per minute or 8,100 per hour; it has a four-in-one document handler and three paper trays with a combined capacity of 4,300 A4 sheets; it has an advanced touch-sensitive screen-based user interface; finally, it has a finishing station that can hot thermal-tape bind or stitch sets, keeping pace with the processor. The operator controls the machine using the colour user interface screen which is touch sensitive and dual language. The interface allows the operator to pre-program jobs, thus minimising downtime, and to retain complex job set-ups including selection of the reproduction quality for individual pages. The document handler includes such features as a recirculating document handler, A3 and A4 stream feeding, a computer forms feeder and a flat platen for book copying. The main processor module houses the main computers that control the copying machine including a 20Mb Winchester disc of which approximately 2Mb is used for the user interface messaging. The paper trays can be switched and reloaded during operation to maintain maximum productivity. The 5090 is the first high volume copier and duplicator to feature a touch sensitive visual display unit for operating the equipment.

This gives an idea of the scale and technical complexity of the sort of products for which the translation department has developed a production procedure.
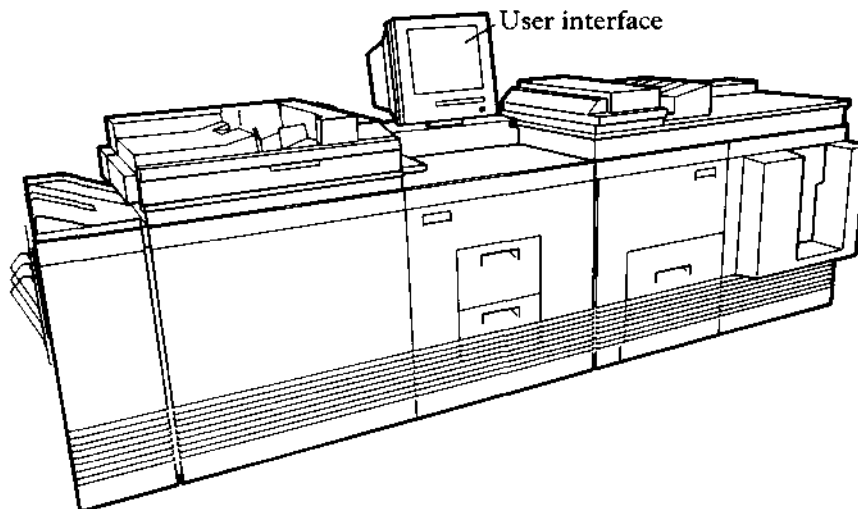

Figure 1. The Xerox 5090 copier and duplicator

FUNCTION OF THE TRANSLATION SERVICE

To understand the translation procedure it is essential to understand the company's objectives, namely: 'That it must be of the highest possible quality to meet customer requirements at the most economic cost and keeping on schedule with the product launch requirements'. Quality is perhaps the Holy Grail for translation. We must seek procedures that ensure the highest quality product so that our customers can obtain the best out of our equipment. The translation process cannot depend on individuals as it would only be as good as the weakest link. It must be a team effort and a rigorous process that works across multiple media outputs and many products: the aim is to create the best corporate image. To stay in business one has to be competitive and strive to be the best and that should apply not only to the product but to all support services. Schedule is the big motivator in today's marketplace. Technology moves so quickly that product advantages do not last and product lives are shorter. The ability to get a product into all of the marketplace quickly is worth millions of pounds of revenue opportunity; it ensures customer loyalty in the future and continuing market share. Translation should never be on the critical path because the lost opportunity outweighs the cost of providing adequate translation capacity.

Some of the problems

The user interface comprises no less than 150 machine operation screens which have to be translated. This is not a trivial task; a simple short word in English may expand to a multi-word expression in another language. All over the user interface screen there are text frames or windows into which words have to be fitted. This is typical of most advanced user interfaces with which translators will have to deal. Expansions of 800 per cent from source to target language are not unknown on the 5090 product. Few software developers are willing to allow for that sort of space; in many cases they do not allow any!

   Even if the translator can fit the words in, there remain the problems of context and adjectival/noun agreements. Some actions are not visible until the translator has viewed the next icon and terms such as 'lighter/ darker' or 'more/less' may not be appropriate to the noun phrase. A number of teams work on the various support delivery media – a general term which covers all customer and service documentation and information whether in hard or soft form – and one of the major problems is to ensure consistency. For instance, the terminology used in the user interface and the operator manual has to be consistent. The solution is to have one process that addresses all delivery media – whether for the customer or the service representative – and distributes the terminology

across all those media. The process should flow from the designer right through to the final documentation and information support media.

XEROX'S STRATEGY

Our strategy can_ be summarised as follows: we use one methodology whatever the final delivery medium. Also, we translate in context so that all the associated and surrounding conditions, whether for documentation or software, are viewed in real time. This means using simulators to represent operation of the product software, since it is not practical to fit the machine on or next to each translator's desk. As a high rate of change is endemic in software environments we aim to translate only the differences between versions. Finally, we maintain tight managerial control over the process especially as regards problem solving. It is a team effort.

Let us look at the translation process and the systems we use to accomplish this (see Figure 2). The key to consistency is terminology development, for which there are two routes: the top route shows the initial pass purely for terminology development using a series of terminological tools after which terminology is reviewed with a translator in the local operating company; the lower route shows that during the course of translation the terminology is constantly assessed in the light of our experience using it, and subsequently enriched.
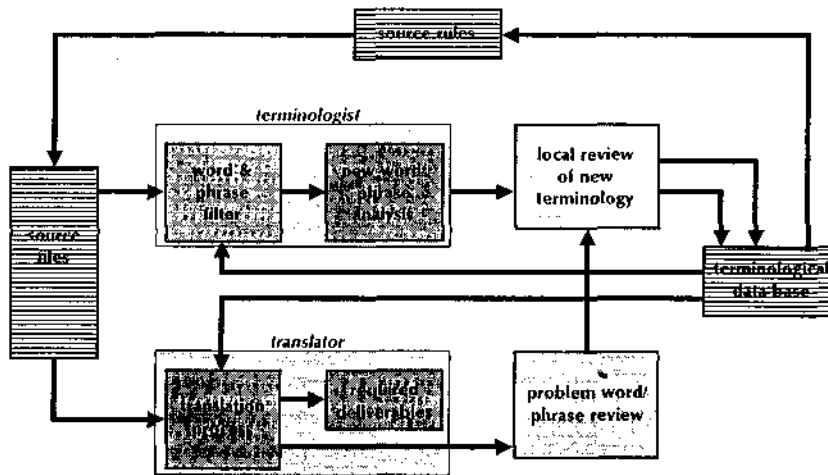

Figure 2. Terminology development

Figure 3 shows the translation process. Firstly, all documents are processed to look for differences; then the non-language elements are treated so that they do not interfere with the next stage of computer-assisted translation; the MT output is post-edited to a fluent language style; the translation is then merged into the end media format; finally, we review the results in context. We have designed the process around the translator who holds the key to the success of the job. The translator makes the judgement and the rest of the process supports that judgement. Problems are flagged, and the translator moves on.

Unresolved issues are then dealt with by the team, going right back to the originator if necessary. Major problems such as whether or not to change the particular software are referred to a management forum since the whole product launch schedule might be affected. It is worth noting that the most stringent review of the intelligibility of the source material occurs at this stage and, since error in the source material may be multiplied over 10 times, all versions are vigorously checked.
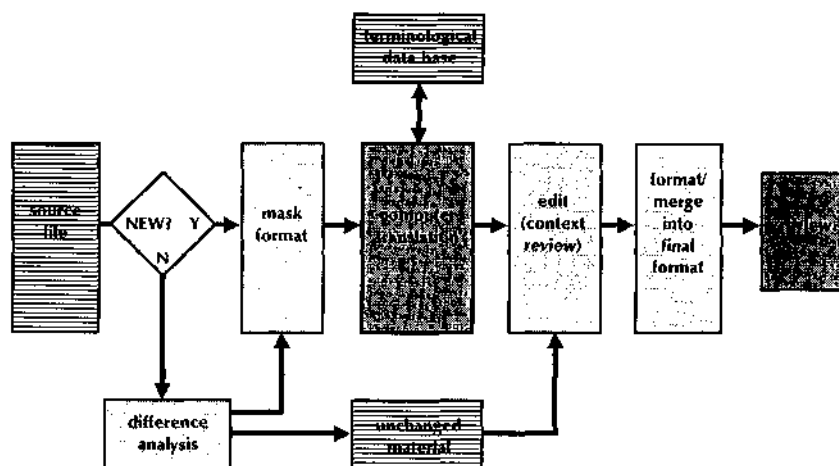

Figure 3. Translation process

Quality assurance begins with terminology control: the job is tackled by a team which includes the originators; we try not to change what we have already got right; editorial checks are carried out during the translation process; final validation is always made by an independent person who compares the translation with the product. Figure 4 shows the complete translation process including the control processes which occur at the main stages of translation.
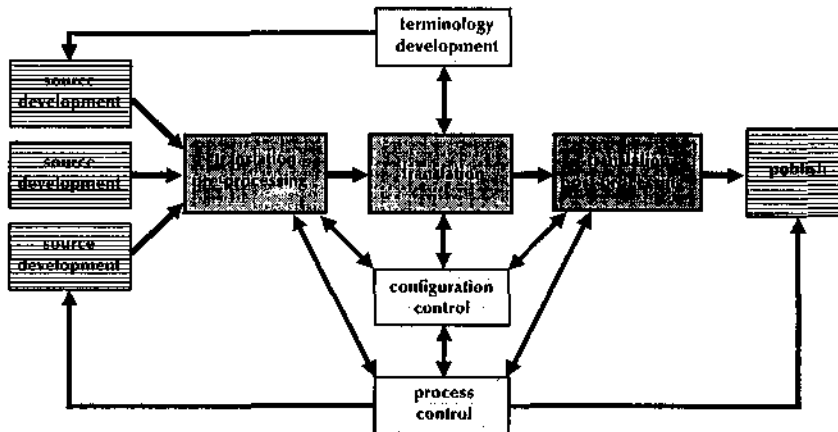
Figure 4. Total translation process

THE METHOD OF TRANSLATION

Translation of software

The translator works at a 1186 Artificial Intelligence workstation running Interlisp D which simulates the user screen for the 5090 copier. Translations from either the Systran or ALPS machine translation systems are sent to the edit window which is sized according to the space available in the source document. The format of the text can be decided at this point. Having decided on the aesthetics, the translated text is then substituted for source text on the screen. In this way the translation is added to the simulation of the user interface ensuring that the fit and function are good. The whole system runs on an Ethernet Local Area Network (see Figure 5) which connects all the devices together. This includes the translator's A/I workstation, the minicomputers running the ALPS software, the mainframes running the Systran software and the file services and printing systems.

Translation of documentation

To translate documentation the translator works at a 6085 workstation, using Viewpoint software which simulates the working environment of the office desktop. It is very similar to the Documentor software but includes networking for mail services and central archive and storage. Most Xerox documentation is written using this type of workstation. By clicking the mouse on the Systran or ALPS icon a document can be sent to one of the MT systems for translation. The translated document can then be returned to the same workstation with the original format retained.
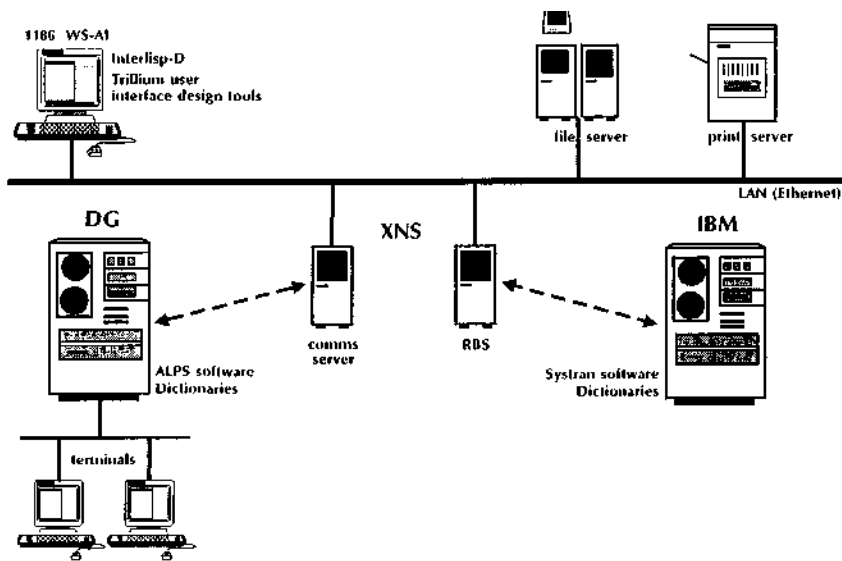
Figure 5. Ethernet Local Area Network - translation of software

Any annotations on the graphics would also be processed. Systran output can be post-edited *in situ.* At present ALPS output is edited in interactive mode and sent to the workstation ready for final production. Other reports such as the Systran not-found word list can also be sent to the workstation. Here again the network which supports the translation of documentation (see Figure 6) is very similar to that used for software translation.

ADDING FREELANCE TRANSLATORS

There was a danger that we were becoming too centralised so we considered how we could introduce freelance translators into the network. This would give us more capacity and also more flexibility for extending the workforce when the load was high. The challenge was to retain the same procedural discipline. We use IBM PC/ATs running ALPS software under Xenix. It is early days yet and we are still working with translators to ascertain the optimum procedure for both parties. Again, the network diagram (see Figure 7) looks very similar to the previous diagrams but this time shows the PC/ATs on the network communicating with the machine translation systems. The jobs/ dictionaries can be prepared before being shipped out to the translators either on floppy disc or via a modem and telephone line.
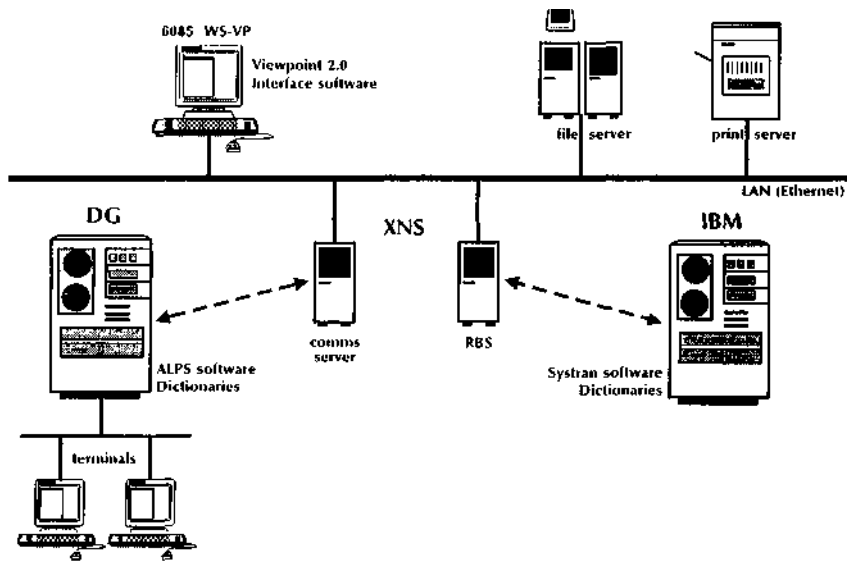
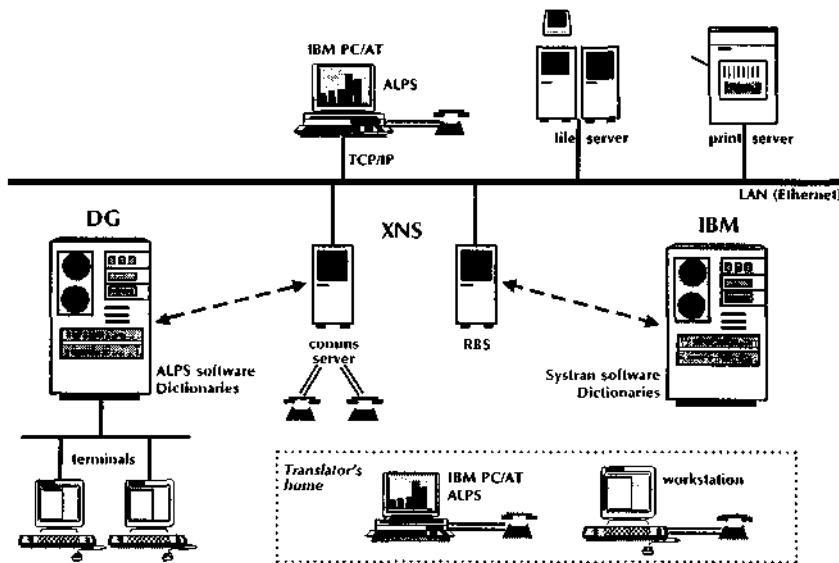**Figure 6. Network for translation of documentation**

Figure 7. IBM PC/AT linked to the network

Thus everyone at Welwyn Garden City shares the same network (see Figure 8) including the machine translation systems, file and archive facilities and printing systems. This network extends to all our major operations in the world, so that the originator may sit in the United States, the translator in the United Kingdom and validation and print may occur in West Germany.
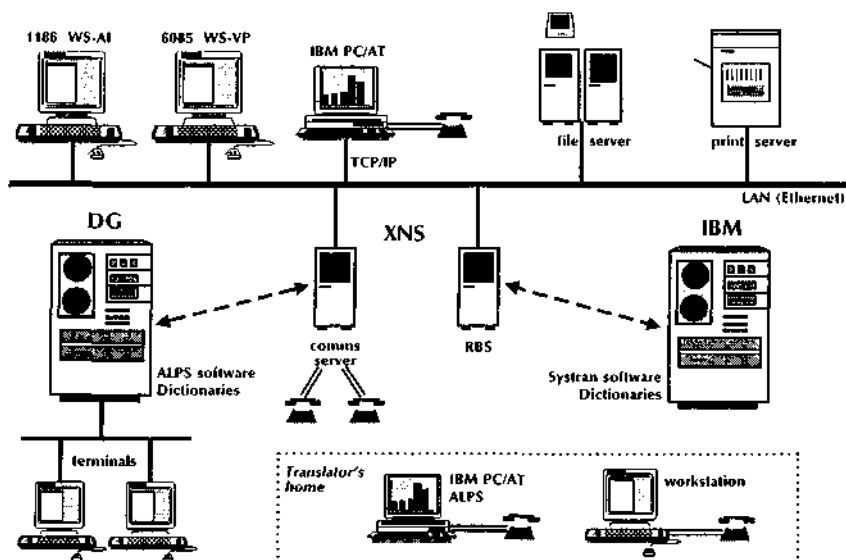

Figure 8. Complete network used by Customer and Service Education

THE BENEFITS

What are the benefits to Xerox? Firstly, the integrated network is based on our own products. The process is flexible in capacity and in its location across the world. It shares the company network through which all mail, office communication and publishing is conducted. You may have noticed that we use two machine translation systems. This is because we believe that they each have their strengths and weaknesses and we have targeted them at different types of translation. We have been long-term users of Systran and have been very happy with the performance when preparing information for our service representatives. As originators we had total control of the process. However, when we considered extending its use to customer materials, the investments in language pairs and dictionaries proved to be a major problem. We also considered that ALPS offered a lower risk especially when handling inputs from multiple

originators, if not the full raw power we were looking for. In addition, there were less integration problems with ALPS and the range of equipment we could use was greater. Their co-operation as a vendor has been high, although it is as yet unclear how new developments at ALPS will affect us.*

CONCLUSION

In summary, we use one methodology, one process which can encompass different MT or MAT systems and which works for different media. We aim to provide the translator with the best possible translation environment and the whole process integrates designer tools, translation tools, publishing tools and all the people that are involved from originator to user. It is aimed at giving the customer the highest possible quality product. Progress comes from making things simple, giving power to the people concerned, in this case the translators, not allowing the process to be compromised, creating a team effort and lastly, ensuring that the customer is king.

We have made a lot of progress in putting translation through one single process addressing whatever output media we are translating. We have integrated very dissimilar hardware and software products into a design, translation and publishing environment, and we still have some way to go. From a translation technology standpoint there is little progress in the industry but we are thinking hard about the design of Mark II!

AUTHOR

Mike Scott, Business Systems Manager, Rank Xerox Ltd, Customer and Service Education, Bessemer Road, Welwyn Garden City, Hertfordshire AL7 1HE.

* Ed. Note: See paper by A. J. Zirkle, p. 11.