[From: *Translating and the Computer*, B.M. Snell (ed.), North-Holland Publishing Company, 1979]

COMPUTER AIDED TRANSLATION

A Business Viewpoint

John S.G. Elliston

Manager, Customer & Service Education (Uxbridge)

Before one starts to look for a particular solution, it is necessary to define the precise needs of the problem. Such is the case with our Company; the solution we are pursuing is tailored to the specific communication needs we have identified and it may well not be the most effective direction for another Company. In order to understand why we have chosen our particular path, it is helpful to explain briefly the Company environment.

THE ENVIRONMENT

Xerox operates in more than 36 Countries spread throughout the world. The task of our Technical Service function is to install and maintain our products, both rented and sold, in each of these Countries. Although the size of operation differs considerably between Countries, the individual functional support needs, in terms of technical data for our Service representatives, is virtually identical. The data is provided in the form of Service Documentation.

This documentation is vital for the field service organisations to be able to do their job. The documentation provides the Service Representative with all the technical data that he needs. It comprises maintenance procedures, technical data, diagnostic procedures and spare parts lists. This type of information must be provided for all products and all configurations. To provide all of this data, we go through the following processes: documentation development, validation, translation, production, distribution and maintenance. An operation of this type is complex enough, the pressure related to accuracy and timeliness for individual locations just adds to the difficulty. Materials are developed prior to launch and "in-field" validation tests in English may well be running concurrently with several translation programmes, in order to enable the staggered National launch needs to be met.

Of the 36 Countries mentioned, only 7 have English as their first language. Even within this group there are sufficient differences within the languages to cause some misunderstanding. A further 14 Countries are obliged to use English text documentation largely because of economics related to the scale of operation. Within this group the ability to speak English varies from very high to very low. The remaining 15 Countries require that all documentation is translated before it can be used in their field environment.

The whole operation is critical and costly and any inefficiency can quickly escalate costs. Consequently, the process needs to be subjected to tight controls.

THE PROBLEM

The problem can be broadly expressed under three headings:

- **Costs**
- **Timeliness or lapse time.**
- **Clarity of communication.**

i) COSTS

The demand on our translation resource grows every year. This demand is related to our increasing product range, refinements to existing products and the normal on-going need to maintain existing documentation. An additional factor is the legal demands placed upon a multinational operation to translate to meet legal requirements. One obvious answer is to increase our resource to handle the growing load. Unfortunately, increasing the translation resource increases our cost base and makes us less competitive. The solution we need must be found in productivity, i.e. using the resources we already have, more efficiently.

ii) TIMELINESS OR LAPSE TIME

Service documentation is developed by our headquarters function either in the US or UK. In either case, it will be originated in English. On average, it will be between 3 to 4 months after the first English version has been validated before a translated manual will be available. (The precise figure will depend upon the complexity of the product and translation workload and prioritisation). This lapse time reduces the possibility to field-test products in non-English speaking environments and consequently, puts a heavy burden on the English speaking Companies, who must now do the majority of the field tests. The question of validation in non English speaking markets is further compromised by localised translation. Manual translation will inevitably be tempered with experience and interpretation. This means one is no longer testing the original, therefore results obtained are invalid. This is even more of a problem if the 'subjective' or the 'interpretative' translation actually improves on the original English version. It is one thing to identify a documentation fault and relate it to either an origination or translation error; it is quite another thing when a problem is resolved by the translation. In the latter case, the translation passes the test and the original English version gets printed complete with fault.

Timeliness is also a key feature of our documentation corrections and update system. At present, extensive delays can result before translated data is available to the field. Again, the reasons are the same, the complexity of the task and prioritisation.

iii) CLARITY OF COMMUNICATION

The two major factors that contribute to ambiguity within our multinational environment are:

- **ambiguity - text must be written in a clear manner.**
- **vocabulary - text should only contain those words that are known to be in the end users vocabulary.**

A commonly expressed opinion is that if a group of 50 translators were given the same sentence to translate, they would produce 50 different versions. A computer given the same sentence will only give one translation. How can we assume that the one output from the Computer is the right translation. I believe that the question directs our attention to the wrong place. The real problem is in the fact that the original sentence was capable of 50 different interpretations. To the producer of Service Documentation, this is frightening. If one sentence is open to so much interpretation, what chance does a Service Representative have when one realises the permutations of a complete book? Obviously, the first problem to tackle is the generation of source material.

Our experience to date has shown that it is extremely difficult to define clarity sufficiently objectively to ensure an author writes clearly. Each writer has his own personalised style. Simply using good grammatical English does not in itself eliminate ambiguity. If a writer has written a piece of text that conforms to grammatical rules, the question as to whether it is ambiguous usually results in, at best, subjective discussion and, at worst, emotive argument.

Secondly, with present techniques, it is not possible to ensure that authors only use those words within the vocabulary of the target population. Our target population spreads across 36 Countries and ranges from 18 to 50 years of age.

Added to this situation, we find that all too often words or phrases have a different meaning when placed in a different context or worse still, in another cultural environment. Recently, whilst visiting the United States, I purchased a coffee from a secretary who looked after the departmental percolator. The price for the coffee was a very modest 10 cents. When I learned of the low cost, I mentioned that it was very cheap. This comment was followed by a rather obvious silence. My colleague later pointed out that it would have been better to suggest that the coffee was 'inexpensive'. The word 'cheap' in the US is usually used in a derogatory sense. Thus, my rather innocent comment was taken as a criticism of the coffee and could have resulted in my having to find an alternative source of coffee.

It is perhaps this type of apparently insignificant interpretation that can so easily result in misunderstanding, or even offence. This is especially risky where we tread the often delicate path of operating across National practices and Customs.

THE NEED

The need is relatively straightforward. Our Company needs a means of communicating technical data, instructions and information to our worldwide Field Service. The method chosen must be acceptable in the business sense, that is the costs incurred must be less than the benefits gained. It must be capable of providing the output communication when and where it is required. Thirdly, it must ensure that the end user can retrieve data accurately and quickly.

Finally, throughout the complete cycle from generation of source language text to translation into target language text, the needs of three categories of end user must be met.

. Personnel whose first language is English.

. Personnel who are obliged to use English but whose first language is not - English.

(This particular requirement adds a third dimension to the discussion of manual vs machine translation, as it automatically forces one to look more closely at the source language).

Personnel (or machines) who are required to translate the English text into the target language.
(These people differ from category two in so far as they will not have the same depth of technical knowledge and understanding).

THE SOLUTION

There was and still is, no instant or obvious solution. The path we have followed has taken us through several potential solutions, each in turn being discarded until we have arrived at our present status.

Perhaps the first approach that we looked at, was one based on the "Caterpilla English" concept. At its simplest it is a limited vocabulary with each word being carefully defined. The target population is then taught to recognise the words rather as one would recognise a symbol, then associate it with the defined meaning. The end user is not taught to pronounce the words, just to recognise them, thus he does not actually learn the source language. This method has been successful in many areas, but did not fit our particular situation. It would be true to say that we rejected the system more on social grounds than on the basis of any real scientific testing. A new Company setting up its operation may well find the system workable, although legislation in some Countries might make even that difficult. In our situation, we were dealing with well established Operating Companies who already translated material to a high standard and a Field Service force, used to having their support documentation in their own language. To switch to a limited English language was seen as a retrograde step and totally unacceptable.

A second solution considered, was the use of a "Command English". This looked a far more likely solution as one could fairly accurately prepare translation for standard command sentences. This would achieve two things. Firstly, a guarantee that the translation is accepted in advance and secondly, a machine can be used to speed up the process. The difficulty that was encountered in this attempt was the constraints placed upon the source language writer. Much of the Service information can be expressed in the directive manner of command English. The problem starts to show when one writes "descriptive statements" or "test objectives" or even statements relating to judgements. In addition, the potential for developing the Command language for use in the areas of training and Customer documentation seems almost zero.

During the period these approaches were being investigated, we also examined some of the claims at that time for existing computer translation systems.These systems by and large claimed to offer unrestricted input translation and seemed promising. Regrettably, these claims seldom lived up to the test and the systems tended to be extremely expensive in development and post edit costs per language.

The system that we are currently using to develop our total translation process is SYSTRAN. Initially, we did some research with uncontrolled input text which resulted in unacceptable output in terms of the post edit effort required.

The dilemma at this stage was that if one used a totally free form of input, the computer translation output required a massive post edit. Conversely, if the source language was written to permit computer aided translation it became unacceptably restrictive to the author or originator. An additional problem with

this tightly controlled input is the acceptance of the user of source language material.

The large post edit task was unacceptable in terms of both cost and job satisfaction. The amount of post edit was such that it took almost as long as it would have taken to translate the whole exercise manually. The morale of a translator in this mode of operation is low. After all, the job is reduced to trying to understand and correct a rather badly written document.

By now it was clear that computer aided translation was achievable but its acceptability was related to the balance between the control of the source language input and the degree of post edit required of the target language output. On the one side, if the constraints placed on the originator are too severe the increased load would cancel the productivity benefit of the system. In addition, one runs into the - real danger of author motivation. On the other side, if one relaxes the input control on the source text translation too much, the post edit function grows to the point that machine productivity is wasted and a similar motivation problem exists, this time for the translator.

The input controls that we have placed upon text origination falls in two main categories:

Vocabulary - It became necessary to ensure that misunderstanding or ambiguity did not arise out of the use of a particular word, or because of the context in which the word was placed.

Writing Rules - Once again, to reduce ambiguity in the source text it was necessary to determine rules to define the required size and construction of sentences, etc.

The vocabulary was developed by combining the work initially done in our US and UK based locations. For example, in the UK location we had developed a vocabulary from ILSAM (International Language for Service and Maintenance). This vocabulary became known as RX Customised Vocabulary. The vocabulary was compared to one that was developed in our US location and from the two sets, we developed our present lists, now known as MCE (Multinational Customised English). This vocabulary is made up of several sub-groups.

Firstly, the basic core group vocabulary consisting of approximately 1000 words. This group forms the basic communication word list. The other groups are to permit the specialist communication within our specific Company environment. They fall under the categories of copy quality terminology, publications terminology, abbreviations, weights, measures, etc. In all, this provides a total vocabulary of under 3000 words.

The next step in the process was to get each target language user to identify their own language equivalents for each word in the MCE vocabulary. As anyone who has tried will know, selecting one foreign language word for one English word is a tough proposition. The important factor is not to simply look for a word for word equivalent, but define one and only one meaning for each English word and then find the target language word or phrase to relate to that precise meaning. For example, the word "replace" is often used to request two quite different actions, e.g.

. Remove part A and replace it with part B.

. Remove part A, adjust part B and then replace part A.

In the first case, we are using the word to mean "exchange" and in the second case to mean "put back". This usually gives little problem to experienced English speaking staff, but does cause problems for those who use English text, but whose first language is not English. It also gives problems to the computer.

Again, for the sake of clarity, each word was defined as a specific part of speech and, if possible, never in more than one category, i.e. "Switch" as a noun and not as a verb. Unfortunately, this was not always possible.

At this point of time in the development cycle, we are gradually being forced to face a simple truth. The computer refuses to understand unless we write clearly and simply. This should not be a lot to ask, in fact it is really what our Service Representatives have always required. Seen in these terms, the project seems reasonable. If we are writing service support documentation with a vocabulary that people are not familiar with, then we are clearly not doing our job effectively.

The same correlation is found between the needs of the end user and computer in terms of writing rules. If sentences are written simply and kept short, then the target population is satisfied. For example, an English technician may have no problem with the following statement:

> "Loosen main motor and Drive shaft and slide back until touching back plate".

This statement demonstrates at least two problems. Firstly, the sentence is too complex. It needs to be written in several short sentences. Secondly, in an attempt to reduce the amount of text the technician must read, we tend to leave out the definite article. Again, anybody familiar or trained on the subject and who speaks English fluently will probably have no problem. The computer unfortunately has neither of these two advantages.

Imagine reading a telegram, (usually written in abbreviated form to save costs), "SHIP SINKS". Does it mean "THE SHIP SINKS" or "SHIP THE SINKS". The difference in meaning by simply moving the position of the definite article is enormous. To overcome these difficulties, the original statement should be written as follows:

> "Loosen the Main motor. Loosen the Drive shaft. Slide both parts until they touch the Back plate".

Summing up, the input to the computer is controlled to the extent that it must be written within the vocabulary of the computer and written in simple short sentences.

One concern that was originally felt by the English speaking service representatives, is that the text that they would be issued with would be written in a form of pidgin English. The example in figure 1 shows this is clearly not the case. Our field test indicated that of the order of 90% of our UK sample found no difference between the ordinary text and the customised text, in terms of usability. Our tests in Sweden, where English is the second language, indicated a 70% response that found the MCE version far easier to understand and use.

CHECKS AND ADJUSTMENTS

CONCISE

WARNING; MALADJUSTMENT OF PLATEN COVER INTERLOCK SWITCH S15 AND INCORRECT OPERATION OF THE MERCURY SWITCH S2 (16.3), CAN EXPOSE OPERATOR TO EXPOSURE FLASH. ENSURE BOTH SWITCHES ARE OPERATING CORRECTLY.

1. (B5,B6) Loosen screws (G) front and rear.

1. Loosen front screws (G). Remove hinge. switch S3. Loosen spacers and screw, that secure rear hinge.

2. Put (B5) 600 T 91030, (B6 onward) 600 T 91197 under platen cover (F). (B5;B6) Latch, then press down on platen cover. Tighten screws (G). Remove 600 T 91030.

3. Latch, then press down on platen cover. Tighten screws (G) and spacers. Remove 600 T 91197. Put back interlock switch S3. Adjust, 16.2A Top cover interlock switch S3.

4. Loosen screw (A). Bias S15 towards the document glass. Ensure actuator arm on S15 clears roller on latch arm (B) by 0.13mm (0.005in). Tighten screw (A).

5. Loosen screws (C).

CUSTOMISED

WARNING: IF SWITCHES S15 AND S2 DO NOT OPERATE CORRECTLY THE OPERATOR WILL BE EXPOSED TO THE DANGER OF 'FLASH'.

1. On B5 machines, loosen front and rear screws (G).

2. On B6 machines and later models, loosen front screws (G). Take off switch S3 (A1, Figure 16 Part 1) and loosen the spacers and screw that hold the rear

3. Put 600 T 91030 (B5) or 600 T 91197 (B6, B7) under platen cover (F).

4. Close and push down on the platen cover (F).

5. On B5 machines, tighten front and rear screws(G). Remove 600 T 91030.

6. On B6/7 machines, tighten screws (G) and spacers. Remove 600 T 91197. Install and adjust switch S3.

7. Loosen screw (A). Move S15 towards the platen glass.

8. Adjust so that there is a 0.13mm (0.005 in) clearance between roller K, (16.1.B) on catch arm (B) and the actuator arm on S15. Tighten screw A, (16.1.A),

9. Loosen screws (C).

FIGURE 1

It is important to stress that the judgement in terms of final acceptability is not that of the originator or translator, but that of the end user, in our case the Service Representative. The judgement is based upon the end user's ability to follow the instructions easily and quickly with no negative impact on job performance.

With the computer programmed and the necessary vocabularies or dictionaries and target language loaded, the system is ready to go.

The source language text is fed into the computer and the target output can be delivered in either hard copy or displayed on a video display unit. The next stage is to post edit the output. This involves identifying errors, analysing them and determining the cause and, if possible, determine solutions to eliminate similar errors in future. These solutions might fall in one of several areas. It may be necessary to add to, or modify the existing dictionaries. It might be necessary to alter the writing rules for future use or it could be that the computer software needs adjustment. Each of these actions has a cumulative effect, gradually taking the total process nearer to the minimum acceptable productivity targets set against the system. It can be seen that in the early stages of implementation of such a system, it is very much a question of "running in" the system.

STATUS

At this point in time, we are extremely optimistic that computer aided translation, using our input controls and based on SYSTRAN, can be used to significantly improve our translation function. We have already shown translation productivity gains of better than 4:1. This level of productivity includes the post edit function related to computer translation. Evidence to date suggests we will improve this level of productivity as we continue to use the system and reap the cumulative effect of software and file improvement. So far our tests have been limited and "off line". The programme that we are working on now is designed to test the total process. This process will involve on line authors' originating the source text using the writing rules and the MCE vocabulary. The text will then be run through the computer and post edited by a qualified translator for the target language. Translation is only part of the total process of developing and implementing a Service Documentation system. As in other systems, there is little to be gained in speeding up part of the process if you leave a bottleneck in another part of the system. For example, there is little gain in spending millions of pounds to build a motorway if all it does is speed up the traffic to the motorway exist and create a traffic jam at the intersection. So with our approach to translation, it is an integrated part of a total system.

Once the post edit stage has been completed, the system will permit further productivity benefits. As all the text, both source and target language, is held in the computer we can electronically file it, update and modify it and print it out. By hooking our translation systems directly into a computerised text editing system, we can automatically select type face, size, etc. and compose the final page on a video screen. This greatly speeds up the total process and eliminates the relatively slow and expensive text creation and composition stages each language has traditionally required.

To date, we have carried out tests with English to French and Spanish translation. Other language pairs will follow, but in each case it is essential to ensure the end user of the target language is involved in the development process. One obvious example where it is essential to gain acceptance from the end user is where you

are selecting one base language that is to be used in more than one Country.
e.g. French - France, Canada, Switzerland, Belgium.
Spanish - Spain, Latin America
Dutch - Holland, Belgium
Portuguese - Portugal, Latin America

In each of the above situations, the variations of both languages in each Country are significant. However, it is possible to gain acceptance for a common vocabulary between Countries by careful selection and discussion.

Exactly how we will finally install the system in terms of function and location is still under development considerations. The diagram (fig. 2) shows the principal activities that we are 'hooking' in to the system. At present, the post edit function is done from hardcopy, as it is an integral part of the input and computer software preparation. Once the system is up and running and post edit becomes purely a translation/editing function, the work could be done remotely or on site, direct on a video display unit.

At this stage, we cannot say for what ranges of application computer aided translation will be suitable. As was said earlier, the system can be used for pure technical communication, where facts are listed for future retrieval. Whether it is possible to extend this into the area of training has yet to be established. Within the definition of training in our Company, we range from pure technical skill-based training to the highly interpretive interpersonal skill training. It seems reasonable to assume that the more straightforward technical training offers the best opportunity to use Computer Aided Translation. However, it must be appreciated that training materials are not written to record data but to enable initial learning. To this end, training material tends to be written in a more personalised style, making use of colloquial expressions and localised examples. Obviously, this type of translation requires a combination of the translator's skills and also those of the Trainer. In short, we are now in the field of interpretation, rather than translation and at this moment in time the computer falls short of that particular target. Already computer aided translation has come a long way, there is every reason to believe it will go still further. Our judgements on its acceptability must be based on realistic performance criteria and not on subjective argument. We are not trying to perfect an automated system that "appreciates" the finer points of a particular language, but a 'tool' to assist us in the functional translation of a specific area of business communications.

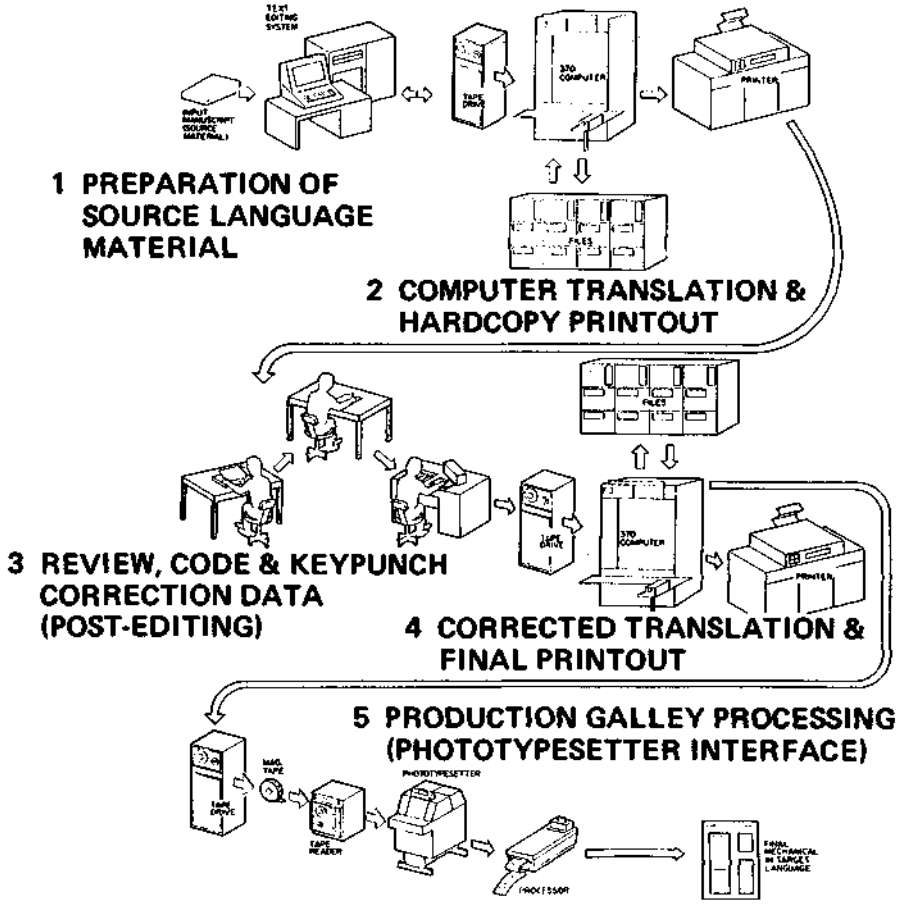
1 PREPARATION OF SOURCE LANGUAGE MATERIAL
2 COMPUTER TRANSLATION & HARDCOPY PRINTOUT
3 REVIEW, CODE & KEYPUNCH CORRECTION DATA (POST-EDITING)
4 CORRECTED TRANSLATION & FINAL PRINTOUT
5 PRODUCTION GALLEY PROCESSING (PHOTOTYPESETTER INTERFACE)

FIGURE 2