[*Terminologie et Traduction*, no.1, 1986]

THE IMPORTANCE OF WORD PROCESSING IN THE USER ENVIRONMENT

Dr. Peter A. Walker
DG V : Commission of the European Communities

Introduction

Some two and a half years ago, I presented a paper to the ASLIB conference on translating and the computer laying down guidelines on the basic requirements for the successful operation of a machine translation system at the Commission. In the interim, we have gained a fair amount of additional experience in this field but I am convinced that a great deal still remains to be done if we are to take full advantage of this relatively new tool.

As a user of the Commission's translation service I will therefore attempt to enlarge on the kind of developments which would be beneficial to the end user in the general modernization of multilingual document processing - including recourse to machine translation - in our institution. In the course of this it will become apparent that word processing is an essential element of the whole package.

There follow four parts

- A specification of requirements
- A comparison of conventional and machine translation (M.T.) from the user's standpoint
- Improving the current M.T. systems
- Development and/or application

which lead to certain conclusions contained in the summary.

1. A specification of requirements

In 1985, with seven working languages or 42 language pairs, the Commission translation services produced over 800,000 pages of translated text. This year, volumes will certainly increase as with the entry of Spain and Portugal, we now have nine official languages or 72 language pairs. Because a number of our translations are used for important political, legal or economic meetings, they have to be of a very high standard of accuracy, whilst still complying with good standards of readability and syntax. Many of the documents are written externally to no common rules of presentation or type-face.

With the basic guidelines of European Community aims largely established, the future trend is for documents to be concerned to an ever greater extent with detail and high technology over an increasing range of subjects. Thus terminology banks and machine translation dictionaries will have to be expanded to take account of these changes. Errors in translation are extremely costly, particularly if they are only discovered at the later stages of agreement on documents or directives. The requirement therefore of a machine translation system for the Commission user is that it shall be:

- convenient
- quick
- accurate
- cheap.

2. A comparison of conventional and machine translation from the user standpoint

On receiving a document in a foreign language, a user must first decide what it is about and how much of it needs to be translated. Few users understand all Community languages and this may cause some unnecessary translations to be made. Then the user has to arrange for the copying of the document, the production of the translation request form, and the assembly of the requisite reference documents in the desired languages. The dossiers go first to the centralized Planning office and then to the translation services. All these stages involve manual transport and all result in delays. The same process applies in reverse for the return of the document, which frequently then arrives as a "brouillon" and has to be retyped and proof-read by the user.

Each of these operations takes time and means that in practice a user must count on at least five days delay between receipt of a one or two page document and the availability of a finished copy in three languages; and approximately four weeks delay for a 25 page document, unless special arrangements are made in advance.

Rapid post-edited machine translation avoids a number of these delays, particularly on the longer documents when net translation rates are approximately four times those of a conventional translation. The machine translation rate is approximately 20 seconds per page.

If the translator/revisor is available and is prepared to work directly onto the visual display unit, then a clean-typed copy of the post-edited translation of 25 pages can be completed to good standards of accuracy and style in a day and a half. The total time between receipt of a document and availability to the user is then reduced to three days as compared with the four weeks cited above. With longer documents of say 40 or more pages, the saving of time is greater.

From the user's standpoint the implied use of a word processor in the translation department (and as importantly in his own) has considerable advantages particularly in the production of texts such as proposals, recommendations, draft directives or policy documents. These are discussed in committee and may finally arrive as the tenth version of the original text. Using word processors the changes can be simply indicated and the new finished copy properly produced with a minimum of effort in both the original and translated versions. Neither the user nor the translator has to check the unchanged parts of the texts.

In toto, the administrator gains in time and convenience if he has the possibility to use M.T. Like most advanced tools, M.T. must be properly handled.

3. Improvements to the current system of machine translation

Six principal improvements are required to the present system. The first three involve hardware; the fifth, software and the sixth involves system control and training.

3.1 Optical character readers

More often than not, a document arriving for machine translation has to be manually typed into the computer system. The first tests with an optical character reader (OCR) by which the manual input typing can be eliminated are very encouraging. English-language texts can for example be input via OCR and corrected by a human operator at rates of up to 70 pages per day as compared to 20 pages per day using traditional methods. For OCR's to have maximum reliability it may be advisable to try to standardise on certain typefaces, number of characters per inch, line lengths, spacings and paragraph spacings. This will present a problem for services receiving much of their documentation from external sources, but this user has always found that when the advantages are explained, our experts are almost always prepared to make a maximum effort to comply with our requirements.

Optical character readers certainly seem to be essential. They must be fast, reliable and be able to cope with as wide a range of material as possible.

3.2 Compatibility of word processors

The Commission has been forced to experiment with many different types of word processor for practical and non-monopoly reasons. These machines are not compatible and in practical terms cannot communicate with each other. It is essential that all user word processors should be able to both send directly and receive directly to and from the main translation systems. Whilst software conversion programs for some machine pairs are available (eg. Olivetti/Wang word processors), the aim should be full compatibility of input and output ports and disc systems for all machines.

3.3 User-friendliness of equipment

MT covers many aspects of word-processing equipment. Excluding those texts entered by OCR, source texts have to be entered by the translation requester typing pool before being submitted by telecommunications to a remote mainframe for Systran processing. Once the raw machine translations are received, post-editing must be carried out, preferably on-screen by the translators themselves, and finally the corrected version must be returned to the requester who may wish to make final amendments.

Given the multilingual environment in which we are involved, such processes not only require sophisticated word-processing software but reliable and user-friendly means of communicating between workstations, printers and peripheral equipment. In this context, user-friendliness means straightforward menu-driven procedures, rapid response times, multilingual screen display and high-quality printing facilities. Users should be able to access all these options without the necessity of loading additional support software by diskette.

3.4 In house modems and telephone lines including storage systems

To avoid the slow process of transmitting documents by hand between the user, the translator and the machine translation system, word processors should be equipped with modems, high quality telephone lines and the necessary storage systems in the circuits to prevent log-jams of transmission so as to enable the machine translation system to function continuously for at least ten hours per day, thus permitting a high return on the large capital investment involved.

3.5 Quality and availability of machine translation

Machine translation has always been criticised for its low quality, and a continuous effort is required to improve the raw output. However experience has shown that when M.T. is consistently employed in a given field, the quality improves as feedback from translators and end users is incorporated into the programs and dictionaries.

Analysis suggests that many of the errors arise from the dubious nature of the input. Each of us has his own peculiarities of literary style, frequently not syntactically correct. The human translator can interpolate these differences and write his text to convey the sense rather than the exact word by word translation: M.T. cannot. The natural question which follows is "Couldn't pre-editing be a more immediate and cost effective development than further syntactic development of the Systran logic?" Whilst this can be applied inside a company with one or two sources, its application to the Commission's business is obviously much more difficult. But for in house authors of documents and translators using M.T., training by means of examples of the results of raw input and pre-edited input, could be an avenue worth exploring. The benefits of pre-editing will obviously become much more apparent when there are a number of target languages as the benefits should reduce the time spent post-editing in all translation services.

However the most urgent matter is to extend the availability of Systran in line with the increased work load expected to follow the entry of Spain and Portugal. New language pairs involving English and French as source languages, and English, French, German, Italian, Dutch, Danish, Spanish and Portuguese as targets, are programmed as a matter of urgency. These pairs will constitute over 80% of the translation load.

3.6 User, translator, and secretarial training

The need for machine translation to be successfully applied within the Commission Services must be explained to senior staff. Such explanation should stress the fact that progress in producing agreement on objectives, policy documents, directives and standards is often retarded by the non-availability of texts in all languages and that the situation is likely to get worse with the increase of the working languages from seven to nine.

In addition it is perhaps possible for the average translator to work from three foreign languages into his own; four at the limit. But to ask him to work from eight is not possible. The effect of the entry of the new countries into the Community will therefore be that a considerable increase in translation staff will be necessary. Senior managers read perhaps two foreign languages, but their reading rates and comprehension in their non-mother language are certainly much lower and limit their decision making capability. Economic factors alone dictate the need for an accurate, speedy and cheap alternative to the conventional translation system.

To develop the application of M.T. users should be encouraged to select documents which are suitable for machine translation, that is where minor errors are not likely to be costly in time or money. Such texts as minutes of meetings, studies and information reports come into this category. To do this, the advantages of machine translation should be fully explained to users.

But the most pressing need is for staff training within the translation services. Many translators express concern about the poor quality of the current machine translations and doubt its advantages, whilst in others there is resistance to learning to use word processors and to edit directly onto them. Still others fear the replacement of artisan skills by machinery. With an expanding requirement for an ever widening range of translated material, these fears do not seem to this user to be justified. Machine translation represents an addition to the traditional skills, competence in which could perhaps merit additional reward as is the practice in other similar professions. If machine translation is really to play an effective part in the improvement of the operational efficiency of the Commission's services in total, it is essential that the interest and goodwill of the translators is maintained and machine translation is seen as an integral part of their activity rather than a separate service and a rival. Notwithstanding this, development of the hardware and systems necessary will have to remain an independent and specialised activity.

4. Development and application

All industrial development is subject to the law of diminishing returns whereby the first steps in any new technology give a very large benefit, but the same research effort later 1n the chain brings a much lower return. Thus there comes a time when development becomes of secondary importance to the next stage - the application of a new process. My personal feeling is that large multilingual users of M.T. are now at this crossroads.

The need for a fully integrated, totally compatible hardware system has been explained. Virtually any word processor can meet the requirements of a multi-language translation user. With the enormous increase in the numbers sold, work processors and most other electronic data handling equipment are subject to rapidly reducing prices and improved reliability. The wisdom of retaining elderly and non compatible equipment rather than the replacement by new totally compatible standardised equipment seems to need to be examined.

It is in this context that it can be argued that a halt should be called to the development of new types of office equipment. If this is so, it is now time to lay down the basic parameters for the electronic input and output of word processors and to buy only machinery which is compatible with that standard. The simplest approach may be to specify an objective which must be achieved by one supplier within a short time span (e.g. compatibility of the output of the OCR with the Systran input; the output from Systran to be compatible with the word processor input, and the output of in house word processors to be suitable for direct input into the Systran system). Then to adopt this as a standard procedure for all machines which may be used in all translation and translation user departments, irrespective of the manufacturer.

A totally different approach is the Commission's Eurotra development scheduled to produce a small-scale pilot demonstration of a new, high quality M.T. system suitable for use between all nine of the Community languages by 1990. This seems a courageous attempt to relaunch the initiative of machine translation, but even when allowance has been made for the fact that Eurotra should be able to build on all the experience gained with Systran and avoid many of the errors and delays encountered, the 27 million ECUs ($24 mio) to be invested between 1985 and 1990 seems to this observer a very small budget and a very small time scale for such an ambitious project. However, even if this does progress, it should at least initially be made compatible with the standard input/output referred to above.

5. Summary

5.1 With the increasing number of languages, machine translation is no longer a toy or a luxury, it is a necessity.

5.2 In the right circumstances M.T. can be time saving and convenient for a user. Proper equipment is required and the documents to which M.T. is applied must be selected with care.

5.3 The immediate aim should be the use of improved hardware, such as OCR's, total compatibility of word processors, and improved machine to machine communication.

5.4 As a method of improving the quality of the raw M.T. output and to reduce the post-editing load, pre-editing of the input material should be tried particularly where there are multilingual target languages for a document.

5.5 The advantages of the system to users should be explained. They should be encouraged to send suitable documents for machine translations. Grading of translations should be considered.

5.6 Staff training, primarily in senior management and the translation services is necessary.

5.7 It is now time to reflect on the balance of effort to be devoted to development on the one hand and application on the other. To speed the application process it is essential that the specification of a common machine electronic output/input code be established as a matter of urgency.

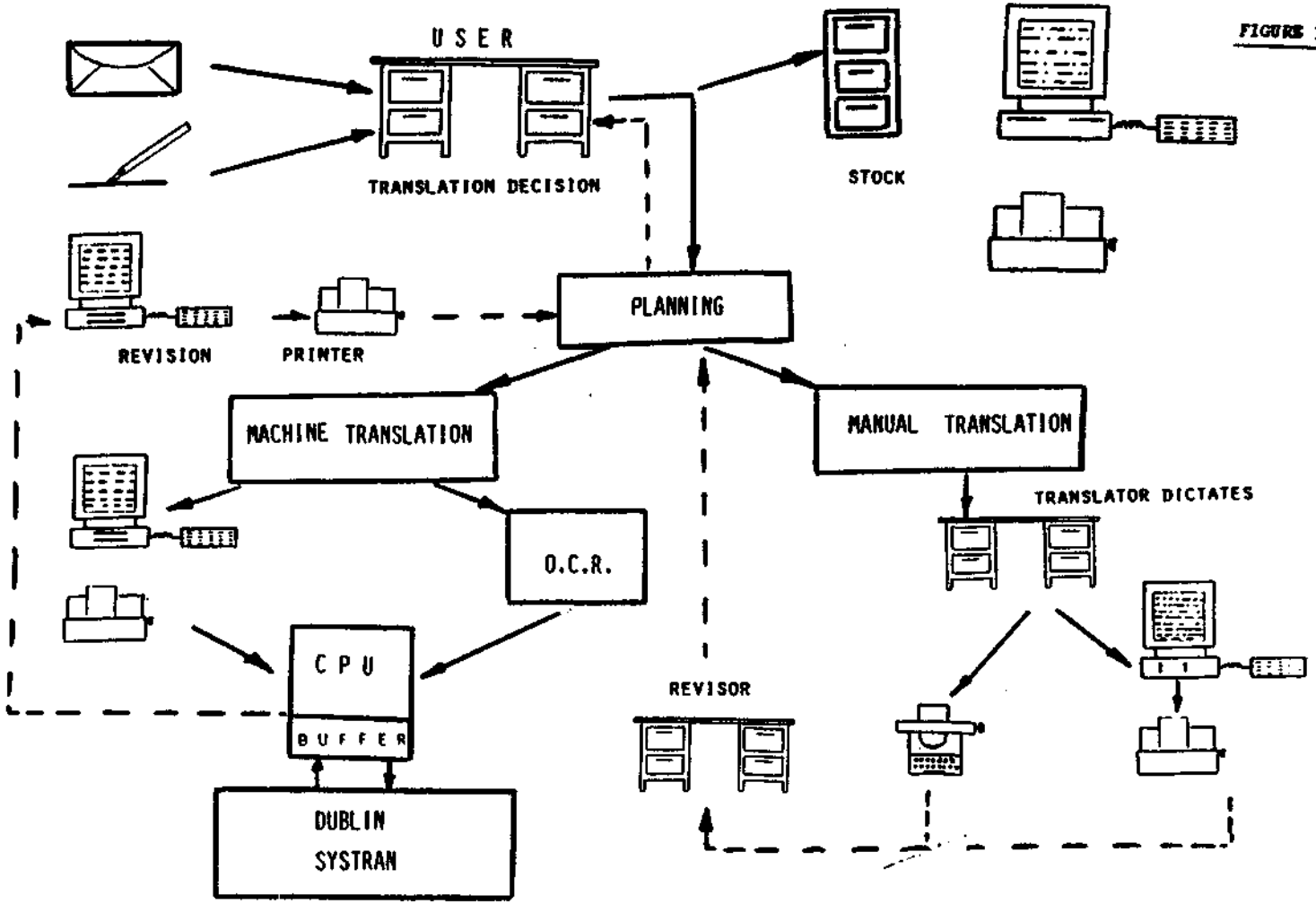
5.8 The key role of the word processor in such a system is evident.

P.A. Walker
Directorate General V for Employment, Social Affairs and Education, Division E/2, Health and Safety.
February 1986

Important: The views expressed are those of the author as a user of M.T., and not necessarily those of the Commission, DG V, or any other of the Directorates General.

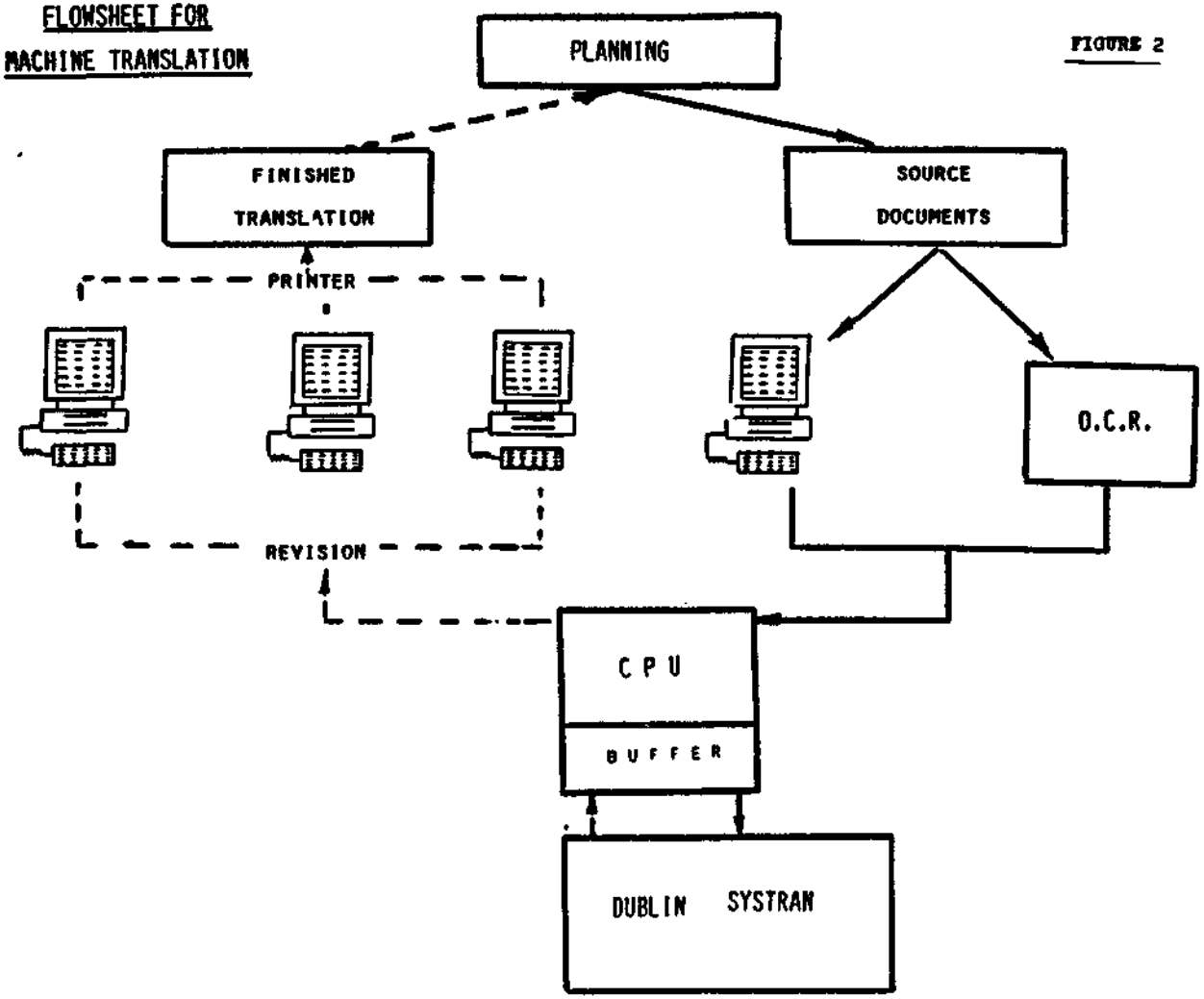
FLOWSHEET FOR COMMISSIONS TRANSLATION SERVICE

FIGURE 1.



FLOWSHEET FOR
MACHINE TRANSLATION

FIGURE 2



ANNEX I

TOTAL OF AUTHORIZED POSTS (LA NOT INCLUDED) COMPARED WITH TOTAL LA AUTHORIZED POSTS (BRUSSELS & LUXEMBOURG TOGETHER)

Years	Total authorized posts (LA not included) (*)	Total LA authorized posts (Brussels & Luxembourg) (*)	TOTAL (*)
1975	6.656	1.120 (**)	7.776
1980	7.500	1.254 (**)	8.754
1981	7.885	916	9.150
1982	8.225	940	9.565
1983	8.473	963	9.840

(*)Permanent and temporary posts

(**) Interpreters included