Machine aids for translators: a review

A. D. Stiegler

Office Automation Consultant, London

Paper presented at the Aslib Technical Translation Group conference and exhibition, London, 20 November 1980

Since translating is an office activity which, like other office activities, consists primarily of processing text, it is instructive to examine the reasons for automating text production. Most of these reasons will be equally applicable to the production of translated texts. We shall then investigate the current developments in machines and micro-chip technology which are applicable to translation. These will include voice recognition and response and optical character recognition equipment amongst others.

YOU MIGHT ASK yourselves why machine aids at all? So I'd like to talk a little about the economic justification for automating office activity. Very briefly if we look at investment in capital equipment in the United States in 1975 we find that the investment in the office was less than a tenth of what it is in the other areas of agriculture and manufacture. If we then look at how that money is spent in the office we find that 86 per cent is labour costs. The office is a very labour-intensive area. (See Fig. l). This is evident from the figure of $2,000 of capital expenditure per annum. (See Fig. 2). If we then look at the costs of one of the primary activities in the office, typing an A4 page, if we look at how that cost has changed over the past 25 or 30 years, we see that it's increasing and in fact today it is around $7.00 per page, simply because it is a labour intensive activity. So what about capital expenditure? What's happening to technology, what's happening to machines? Well, everybody knows about chips by now, I suppose, having seen the programmes on the BBC and on ITV and probably on all the other channels in the world and we know that technology costs are dropping and they are dropping at a very considerable rate. For example the costs of logic, the actual processing of the data that you've got, the cost of that logic is going to drop by 1985 by 95 per cent according to these figures. According to more recent figures that I have in fact it has dropped by 95 per cent today and it will probably drop by quite a bit more by 1985.

So that is why it will probably be cheaper to do it with machinery than it will be to do it with labour. Having said that, what are we doing with machinery? What we're doing with it is typing and you might well ask yourself what's wrong with a typewriter? The primary thing that's wrong with a typewriter is that it has no memory. It cannot remember what you keyed into it yesterday and the entire dialogue that you have with

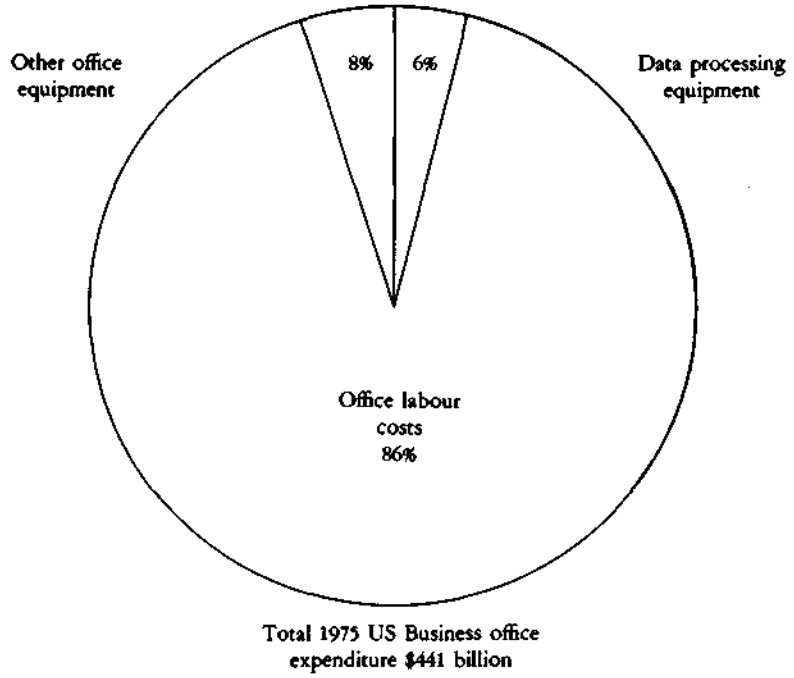

FIG. 1 *Labour versus equipment*

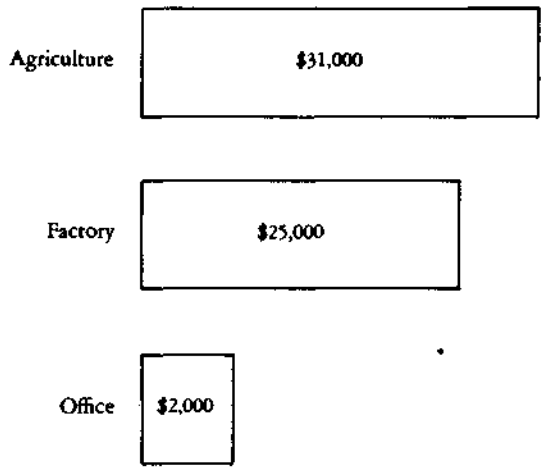

FIG. 2 *US investment per employee 1975*

a typewriter is conducted on a piece of paper. It's hard copy. And as a consequence it's very hard to manipulate it. So if you've made a mistake or if you've decided to put a paragraph in a different place or if you've decided to change something in some way, you can't do it very easily with a typewriter. So what do you want to do? You want to use a word processor. Well, what's a word processor? A word processor is a device which assists in text preparation and production and it has reproduction, storage and retrieval as well as transmission potential, transmission potential being the ability to send, like a Telex, a document from one place to another along a telephone line. So that's what a modern word processor is and what makes it so much better than a typewriter? Typically it has memory to record the typed information and here's an example of difficulty of hard copy dialogue and one of them is: I'll never be able to change that word ''type' to 'typed' because that's what it should be, until I make up this slide again because I haven't preserved that information electronically. So it has memory to record typed information and typically it has a visual display unit with the operator able to see the corrections being made as they are being made.

The whole activity of word processing can be summed up in five headings: it's the sum of all the activities involved in the origination, in the preparation in the production and reproduction, in the storage and retrieval and in the transmission of texts. That's what word processing is all about.

So now we know what word processing is, we know a little bit about why we want to buy a word processor or might think about buying a word processor. And so what should we do when we buy one? The first thing that you should do is examine why you want to buy it and you have to define the contribution that you require of this device. You have to ask yourself: what is it that this is going to give me that I haven't already got? And if what it's going to give you is not sufficient to justify the cost—and the cost of these things is fairly high, make no mistake about it. It's not like buying an electric typewriter. An electric typewriter costs you about £2,000 and one of these things will cost you the minimum of about £8,000. So you have to define what the contribution is so that you can tell which piece of equipment to buy. The next thing you have to do is to determine how to achieve that contribution, how to obtain the benefit that you have told yourself you are going to get by introducing word processing.

The first thing you have to do is analyse existing work loads. For example, in the translation field if your existing work load entails a large amount of report writing and therefore editing, that will make up the bulk of your work, you may find that in analysing that work there are places where you can use standard text or quasi standard text, so you have to analyse that to find out exactly what it is that you are actually doing. You have to draw a plan for revised work loads because what's going to happen is that now, in a typical office—and this is incidentally a general discussion that is not necessarily specific to translation. This is true of any office activity. In any office, what's going to happen with the introduction of word processing equipment is that the work flows will change. The places that work goes to and from will change. And the final thing that you have to do is to get a commitment of management if you don't happen to be your own boss. You also have to get the commitment of the people who are going to use it.

Some of the ways in which you are going to be able to avoid mistakes is to look

at the word processor as a system and not as a collection of individual features. The reason I say this is because when you go to a word processing exhibition, and there is one at Wembley, for example, every year and there's the Hanover fair in Germany and SICOB in Paris which all exhibit word processing equipment, you will be blinded by science. You'll walk in and you'll see these fantastic machines and people will say 'My machine has a feature to ring a bell every time you want to do something or whatever and its got bells and whistles on it which no other machine has'. You can't look at a word processor that way. You have to look at it as a system within itself and also as part of the larger system which is your office. So it has to be viewed in its context and if you feel there is a feature you need, and for translators for example, one such feature might be the ability to display diacritical marks on the screen, have it demonstrated. Don't believe anything that a manufacturer tells you! The next thing you should do, is to test the equipment with samples of your own work so that you can see it actually performing the kinds of jobs which you wish to perform.

A typical mistake which people make in selecting equipment is that equipment is rejected because it doesn't have a particular feature. This could be a mistake because there are certain bits and bobs of equipment around which you think may not have a particular feature which you need but if you find one that does have that feature, it may not have a lot of other features that you really need and therefore you might select it along with this particular feature and make a mistake.

Now one of the problems with word processing and the area of office automation lies in the fact that in most instances, any word processor is better than no word processor because it is going to give you the things that you are looking for. It's going to give you easier editing, it's going to give you, if you're not your own typist, faster turn around from the typist, it's going to give you higher quality output from the point of view of the number of mistakes in it and so on. It's going to give you less requirement for proof reading and all the great benefits one achieves from word processing are going to be achieved whatever word processor you use. But for *your* application, some word processors are better than others and your application over here may not be the same as that person's application over there. He may have a totally different problem from yours. The other problem is that when you do find a feature that the manufacturer has demonstrated, the manufacturer's demonstration may not be understood by you. You may not understand exactly what that feature he's talking about does, and you won't understand it until such time as you start using the equipment yourself.

I've mentioned what word processing offers. It offers improved productivity, that is to say you'll be able to get more work done with the same number of people or the same amount of work done with fewer people typing it or even editing it. The next thing it does is to reduce the amount of time that it takes for the typist to do it. Typically when typing on a word processor when you put the original in, the amount saved is not more than 10 per cent to 15 per cent. Some people I know have claimed higher figures than that for the original typing work. It's when you come to the corrections that the turn around time becomes vastly reduced. You may have forgotten a paragraph or you may have put a paragraph in where you didn't really want it, you have to put it in somewhere else, and it requires a typist retyping two, three or maybe four pages before she gets to a break where everything fits, and then

everything must be proof-read. On a word processor, this doesn't happen. The turn around times both for the typing and for your proof reading are very greatly reduced, and you get better quality. You have fewer mistakes, typically, in a document produced on a word processor than in one produced on a typewriter. If you have to send out two or three copies of something, they can all be originals and it doesn't cost you anything more than just the paper and a little bit of time on the machine to produce three, four or five originals. The thing you mustn'’t forget is the reduction in author time.

Finally, it turns out that in most cases, typists like it. In most installations where word processing equipment has been introduced, typists actually prefer it to their typewriters, because it's a lot easier from their point of view. One of the things for example that a typist does is that when she's typing a line towards the end of the line her typing speed actually slows down because she's anticipating the end of the line and a similar thing happens when she gets to the end of a page. On a word processor this doesn't happen because the equipment automatically allows for the margin and in most cases pages don't really mean anything to a word processor because they regard text as documents. Indeed in these terms a page is difficult to define.

What sorts of word processors are there available today? The first sort that I'd like to talk about is what are called stand-alone systems. A stand-alone system is typified by a configuration such as this. As with any other word processor it has some sort of input and display device for keyboard, typically a television-like screen and a typewriter keyboard underneath it. Unfortunately most of the typewriter keyboards tend to be in the normal typewriter layout which is called 'Qwerty' by the first six characters of the top line of alphabetics, which of course is different in France and different in Germany but we're stuck with that, unfortunately, because of the people who invented the typewriter. It has logic to assist the typist with the typing process. It also has some type of storage and clearly it has a hard copy output device—a printer. That is a typical stand-alone word processor. You might find optional extra features on a stand-alone word processor—you might find an extra disc drive, for example. A floppy disc today can hold something in the neighbourhood of a half million characters and you might find you have two such units so that you could hold a million characters for immediate access. The floppy disc is a magnetic disc about the size of an old 45 rpm record. Now you will find with stand-alone equipment that you can only access 1, 2 or maybe 3 or 4 such discs and therefore immediate access at any one given time is to something in the neighbourhood of 2 million characters, so an extra disc drive can be a very advantageous thing to have. Another optional extra is the communications interface. Another might be an output to a photo typesetting device and finally you might have some kind of calculation mode to assist you in typing columns of figures.

Stand-alone systems tend to meet the needs of a small office. You may find that freelance translators, for example, would be more inclined towards a stand-alone system than they would to a large shared facility system, which I'll talk about next. And they are also good for people who just want to try word processing to see if and how it can help them. Now I'll talk about shared facility systems. The facilities of the word processor are those things we've just talked about—the storage, that is the discs, the logic—both the hardware and software, that is the technology and the program

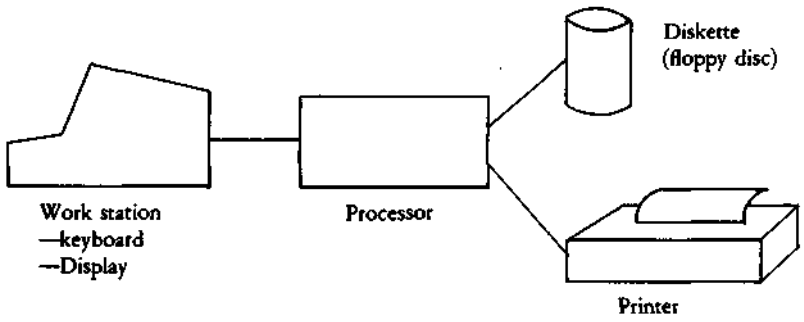

FIG. 3 *Stand-alone system—basic configuration*

to assist the typist, the printers and the communications interfaces. All of these things can be shared by a given machine. One typical example, you might have a set of work stations out here, each of which is being used by an individual typist, a central processor which is driving the work stations, ie it has the logic in it to assist the typist, and it uses a file management sub-system and a printing sub-system and this is a sort

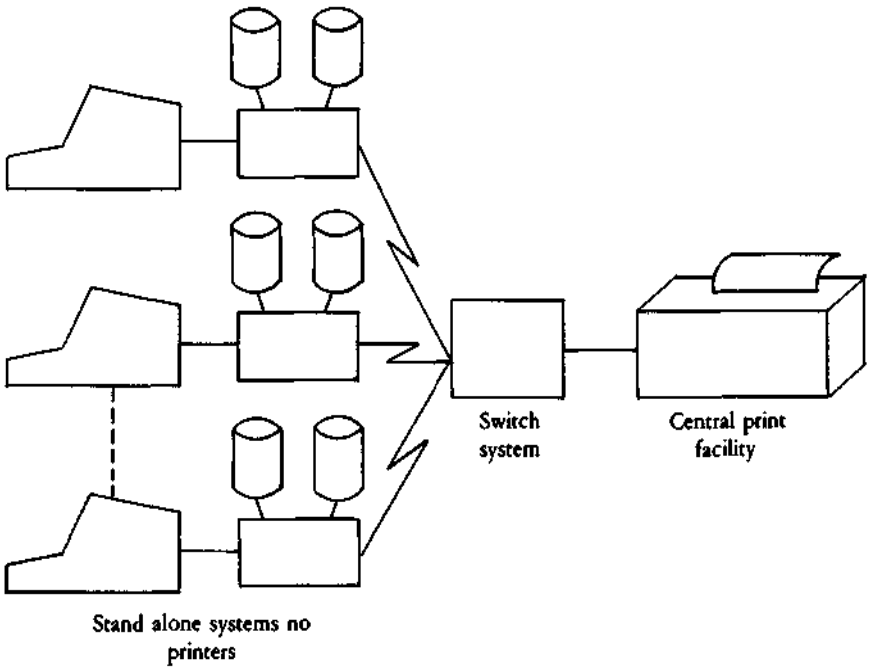

FIG 4 *Shared printers and communications sub system*

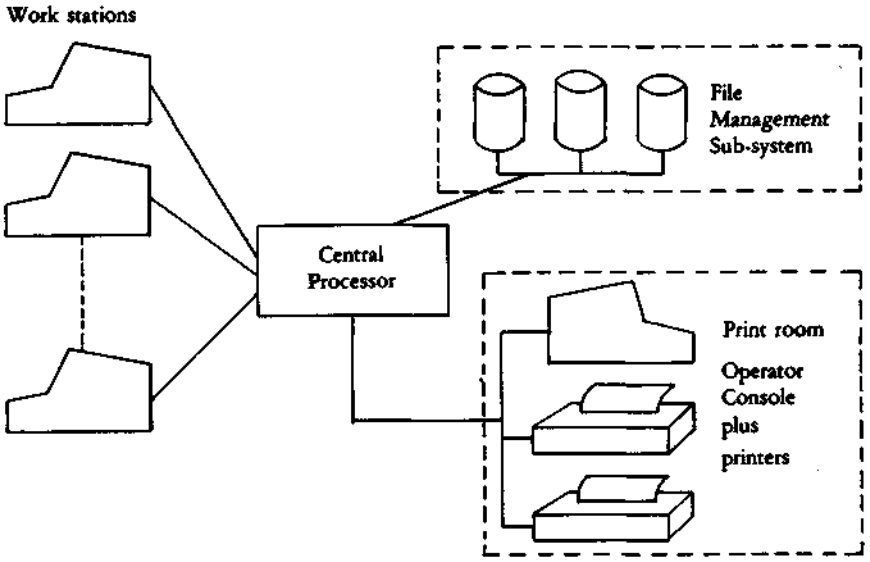

FIG. 5 *Shared logic/typing pool system*

of typing pool environment of word processor, although having said that, one of these work stations could possibly be on someone's desk somewhere else, outside the typing pool area, since there is no necessity for them all to be in one place. So fundamentally that is the kind of system that is referred to as shared facility. There are other examples—see Figs. 3-7.

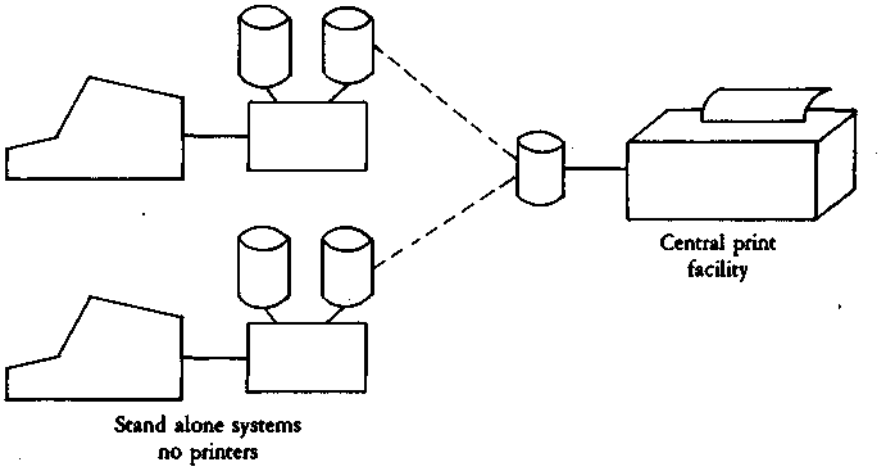

FIG. 6 *Shared printer system (text for printing transferred by manual transportation of the storage media)*

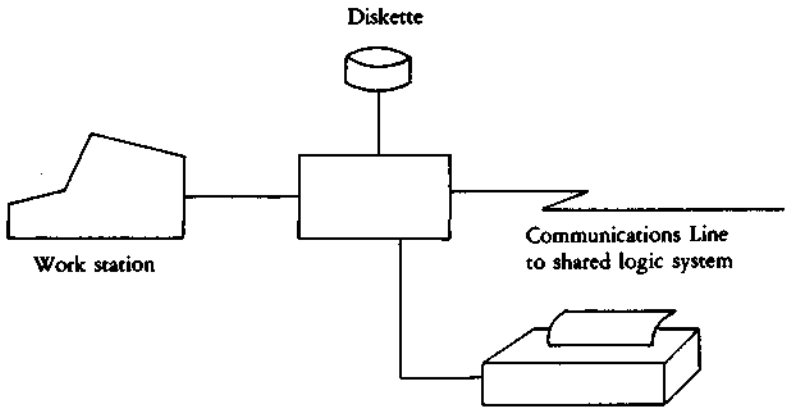

FIG. 7 *Remote outstation*

	Range	*typical good figures*
Stand-alone	0.8-2.2	1.3-1.8
Shared logic	1.2-4.0	2.0-3.0

FIG. 8 *Word processing—productivity*

We've talked about productivity gains and gains in the amount of time and the amount of work that can be performed. What sort of gains would you expect from word processing equipment? A shared facility system—a shared logic system as it's sometimes called—is better than a stand-alone system because it alleviates the problem of housekeeping. As I said before the discs that you have contain a fixed amount of information, say, half a million characters. If you've got more than that number of characters, and most of you probably have, most people would say that's about 100-200 pages—a page is difficult to define. However they have a limited amount of information, whereas a shared facility system, sharing that disc sub-system, might have hundreds of millions of characters of information accessible right now from any one of the terminals and therefore it alleviates the housekeeping problem of taking care of those discs and therefore you would expect the productivity to be greater than you could expect in a stand-alone system. The reason that the number in Fig. 8 (80 per cent) is lower—that is to say with the same number of people and word processors, I got 80 per cent of the amount of work done than I got done before—is because of the housekeeping problem, the problem of floppy discs and the filing of them.

Now we've talked about word processors. What else is there around that might help translators? (See Fig. 9). The first thing that I have listed in Fig. 9 is optical character readers. There are machines around today that can read typewritten text and store it electronically somewhere, normally magnetically. The use of these things is fairly obvious in as much as if you have a typewritten script or a typewritten document

* Optical Character Readers
* Video Disks
* 'Bubble' Memory
* Voice Response
* Speech Recognition
* Electronic 'Phrase Books'

FIG. 9 *Other machine aids*

and you wish to put it on to your word processor, it's much easier to get it on by an optical character reading device than it would be to re-key it.

Another item is a video disc which is a disc which essentially stores images and it stores images in much the same form as a television picture and these images can be re-displayed on a television-type screen. The recording mechanism of the video disc is actually a laser and it can store millions and millions of pages on one very small gramophone record size disc; it could store complete dictionaries that you might need for translating more technical items for example.

Another feature again along the same lines of storing information, storing and retrieving information like dictionaries, is the advent of what is called bubble memory which is a bubble of magnetism floating around in a silicon chip and it can be used, generated, retrieved and saved in this little tiny chip. They come about 7½ cm X 7½ cm and there may be 10,000 characters of information on a chip about that size. Voice response units are devices which respond with a human voice to an enquiry on a machine, these could be more useful to interpreters, perhaps than to translators because if you get into some technical terms which again, you may not necessarily be familiar with, the voice response unit could give you the appropriate pronunciation of that term.

Then there's speech recognition. Speech recognition is an interesting affair because today what we recognize in machines is what's called discrete speech, which is to say you have approximately 20 or 30 or maybe 50 or 100 words, that is words or phrases, that the machine recognizes and can translate into an electronic form. This could be very interesting if we ever get to recognizing what is called continuous speech, translated into some kind of electronic form. This actually has been done in the United States, I understand, but one of the problems is the amount of time and the amount of processor capability that it takes up. One can read a sentence into the machine at normal reading pace and the machine will actually interpret that sentence and put it into magnetic storage in a machine readable format. It takes 75 hours to do it at present, but constant progress is being made.

Finally, what about electronic phrase books? How many of you have seen the adverts recently in the *Sunday Times* magazine for Sharp electronic phrase books? Essentially you key in a word in a foreign language or in English and it displays on a screen the same word in English or the foreign language, according to which way

you are doing it. They're really for tourists fundamentally and I mention them here because the idea is that you can take a calculator-size device and simply by removing one element of that device and putting in another element, you can translate from say English to French in one instance and English to German in another instance. Well, that's interesting but what good is that to a translator? Conceivably you can put in a technical dictionary for civil engineering because you have been translating civil engineering documents and a couple of minutes later you can pull that module out and insert the module for computers because it involves quite a different technical vocabulary to that of civil engineering. Therefore I think that these things can be of use even to the professional translator providing the manufacturers of them—people like Sharp and Nixdorf and Texas Instruments—feel that there is enough requirement for them and again it goes back to the question of the storing and retrieving of information and dictionary look-ups.