

Machine translation in a large organisation

Jean Datta

*United Nations Industrial Development Organisation**

You may be able to swallow a gnat whole, but the only way to eat an elephant is bite by bite. Streamlining the manual operations of an organisation with well-established procedures and a staff experienced in routines by the introduction of computerisation is not unlike eating an elephant. Entrenched procedures die hard and people do not always thank you for disrupting their routines, so the task is best approached bite by bite.

The introduction of computerisation into the existing long-established procedures of language operations represents a challenge to those procedures. Each application introduced brings another wave of challenge, and until the necessary adjustments have been made at each stage, a new wave of challenge coming too soon would leave a language service floundering in deep water. If this is true of less 'invasive' operations such as terminology, then it is doubly true of machine translation (MT), which is the most radical of all the applications available today to assist in language service operations.

In order to analyse this challenge, one must first look at language services as a subsystem within the system represented by the organisation as a whole. The way these services operate is strongly influenced not only by the factors they have to live with on the input side, but also by the demands they have to take into account on the output side. A language service may tend to look upon itself as a self-contained system, operating in the 'best' way to produce the 'best' output. The users, who produce the input and receive the output, are something of an abstraction. They

*The views expressed in this paper are those of the author and do not necessarily reflect the views of the United Nations Industrial Development Organisation.

become real only to the extent that they press for faster service.

Users become very impatient with the slowness of language services, and may find machine translation an attractive alternative. The job that now takes weeks to get done could be fed into a computer, which would return it neatly typed in the output language in a few minutes. The fact is that as long as language services are looked upon as a self-contained system, it may be difficult to make effective use of machine translation, since adaptations to the procedures at the input and output interfaces will be decisive for the acceptability of results. If an effort is made to graft MT directly onto traditional manual procedures, the prospects for success are at best uncertain.

It is therefore essential that, before MT is introduced, the environment, and especially the interfaces on the input and output sides between language services and the organisation at large, should be carefully prepared. This means doing three things:

1. getting the input interface under control;
2. tailoring the output interface to users' real needs;
3. building computer literacy among language staff.

As to the input side, it must first be said that only texts that call a spade a spade are appropriate candidates for machine translation. A text that calls a spade a hippopotamus and does so in many convoluted phrases will be certain to lead the machine up the garden path. As an antidote, controlled language can go a long way towards giving the computer the type of structured task it deals with best. Where an organisation has a staff of professional writers preparing fairly routine texts, such as technical manuals, this is achieved relatively easily, since the writers can be instructed to apply strict guidelines that will make it possible to give the machine a running head start. Where texts sent for translation always go first to editing, there is also some chance of achieving controlled language, although editors may be operating under constraints that prevent them from entirely recasting a text: they may not have time, or the author may squawk if his priceless prose is dismembered, even in the interests of automation. Where much of the translation work originates outside the organisation or is not normally subjected to full editing, the obstacles to achieving controlled input seem practically insuperable.

Even so, there is much that can be done. Staff engaged in drafting can be trained, and the guidelines laid down can be enforced by sending back texts that do not conform. Upstream editing can be used to control input language by sending editors into substantive areas of the organisation, rather as missionaries, to lead those preparing texts along the straight and narrow path of controlled language. These approaches might have some prospect of success if drafters could be motivated to use short, clearly structured sentences, and to refrain from seeking out esoteric terminology.

Of course, trainers and upstream editors are public relations agents, but it is unlikely that their influence alone will be adequate to transform the well-known sow's ear.

Rome was not built in a day, and turning around drafting habits takes time. If top management made clear its insistence on the undertaking, that would be a good start. Language services managers might arrange briefings with substantive officers involved in text preparation to win them over to the idea. Submitting officers should be given feedback about the quality of the texts they submit, to the point of rejection slips if necessary. Those who draft a great deal should be given to understand that they are required to conform to drafting guidelines and their ability to do so should be a factor taken into account in their recruitment. If no opportunity is lost to bring influence to bear in favour of controlling input language, then over a period of time, it should be possible to make some headway, and as one can assume that machine translation systems will at the same time have made considerable strides in the direction of disambiguation, contextual analysis and so forth, there is reason to hope that even organisations plagued by a poor quality of input language at the outset can raise the quality to an adequate level, at least in circumscribed factual areas not requiring feeding to the computer of the sum total of all world knowledge and rules of common sense. In the meantime, conventional translation operations can benefit from the improvements in the quality of input texts by speeding up their service.

As to the output side, traditionally language services have tended to aim at a 'perfect' job, without asking too many questions about the use to which the text was going to be put. However, a perfect job takes a lot of time to produce. There are texts that need to be translated but that need not necessarily be polished to a high gloss. What percentage of an organisation's translation work this accounts for will depend very much on its specific language environment. In some organisations, for example, most texts will be drafted in the lingua franca that everyone understands, and these will be translated for formal purposes such as conferences or publicity operations. In such an organisation, the amount of work needing to be done for information only and not requiring gloss may be relatively small.

Be that as it may, it can be assumed that in most organisations there would be a place for procedures allowing for more flexibility in the processing of texts for translation, in the interests of providing service more closely tailored to the needs of the 'customer'. We saw on the input side how rather radical changes in procedures relating to text preparation could, in the long run, make it possible to control input language so as to take fuller advantage of machine translation. On the output side, too, changes in procedures, in the sense of establishing formal and informal categories of translation, could position an organisation to make effective use of machine translation through rapid post-editing later on, while in the meantime

speeding the flow of conventional translation services.

It is essential that adjustments to the output interface should be made in consultation with users of translation services. Otherwise, the result might simply be a degradation in the quality of service without any improvement in the congruence of service delivered with the users' real needs. If users were simply told that they could choose an informal translation and get it more quickly, they might opt for speed without thinking of possible negative consequences. Therefore, in order to ensure that adjustments made were optimum, it would be necessary to establish some structure such as a working group or task force that would bring together representatives of substantive areas of the organisation that made use of translation services, and representatives of the language services. This mechanism, like those described in connection with the input side, would serve an integrative and public relations function, increasing the permeability of the interface between the language services subsystem and the organisational environment.

The task of such a group would consist of evaluating the categories of work subject to translation and determining the extent to which each could justifiably be subjected to less than the full traditional services. In this connection, there should be extensive discussion of the problems and risks inherent in curtailment of services. Agreement should be reached on the extent to which curtailment of service would be acceptable in the case of various categories of text. On the basis of this understanding, the language services could start a pilot operation that might involve such things as 'one-pass' translation, with no research done on terminological or linguistic problems, possibly even no re-reading of the typescript, no revision by senior staff and no clean-copy typing. Such no-frills translations would be clearly labelled, and the submitting officer who agreed to accept this type of output would also agree to accept its inadequacies, for the sake of speedy response by the language services.

Such a pilot operation would not necessarily involve machine translation, but it could pave the way for its introduction by drawing the framework for the application of the once-over-lightly or rapid post-editing, which would preserve to the maximum the benefit gained through the rapidity of MT. Once the value to the organisation of an informal translation category had been established and it had been determined what volume of work was likely to be subject to this procedure, trials of machine translation systems could be undertaken to give translators an opportunity to gain expertise in the special techniques involved in rapid post-editing and to evaluate the real utility of MT for the organisation.

It is not enough to look at the adaptations to the environment that could be made to ensure the feasibility of using machine translation to produce output of an adequate standard. At the present state of the art, the machine output will in most cases be useless without the active collaboration of

experienced human translators who have the mental flexibility to compensate for the machine's shortcomings in this area. The collaboration between the human and the machine will be fruitful only if the human side of the equation is proficient in interaction with the computer. However, in many organisations the language services are virgin ground, and most of the staff have never had their hands on a terminal.

Therefore, the importance of gradually introducing to computers staff not accustomed to working with them must be borne in mind. The best approach will be to start with applications that require the lowest level of participation by the staff concerned and the least radical adjustment of long-established procedures, and to proceed to those that can only be manipulated to full effect by computer-literate staff. Thus, for example, while word processing is a simple, basic office automation function that any secretary must be proficient in, for professional language staff in many organisations it represents a radical departure from the techniques to which they are accustomed and the skills that they have learned in some cases through years of practice. Obviously, then, word processing is not the best application to start with in such an organisation.

Another factor that needs to be borne very much in mind is that the first applications introduced, while being simple to use, should at the same time be looked upon by the recipients of the system as being evidently useful. If it is difficult for the users to perceive that an application helps them, then they will resist its introduction, and the general progress of computerisation will be seriously impeded. An example of such an application would be MT, even if the output were post-edited on paper printouts. Revision skills, and in particular a sense of what to change and how much, are not easily or quickly come by, and the fact is that some revisers never manage to acquire this sense and continue to rewrite translations passed to them for revision until the day they retire. When a novice in post-editing, however experienced in translation, is confronted with an MT output text that is in places rather rough, he or she may not find it easy to see how MT is helping, since the natural instinct will be to set about rewriting the text. Time and guidance will be needed, during a lead-in period, to develop post-editing skills, so it is advisable not to put MT out in front in any computerisation scheme.

The question, then, is just where to start with the best prospects for success. For example, the first steps might be the introduction of user-friendly databases for document-referencing and terminology, planning them carefully to ensure that, through them, computerisation put its best foot forward in the language service into which it was being introduced. The horizon should not be too short, so that time could be allowed for each new system introduced to be digested before a further challenge came in the form of another new system. A suitable horizon for the computerisation process might be five years.

The technologies relevant to language operations are developing rapidly, so it is difficult to foresee what real possibilities will be offering themselves in five years' time. Will voice recognition have made typing pools a thing of the past? Will machine translation have made human translators a thing of the past? Will image processing and optical disks or other technologies have solved all the problems of storage of archives of multilingual documents? The answers to the questions are uncertain, but what is certain is that if the language service environment is not prepared, the organisation will not be in a position to avail itself of technologies as they come within reach. Any timetable proposed for computerisation over this period is very likely to be overtaken by events, but there are certain parameters, both technical and organisational, that indicate how to set about the undertaking (see Figure 1).

Just as procedures at the input and output interfaces between the language services subsystem and its organisational environment cannot be changed in a day, so those within the language services will need time to evolve. There are opportunities for computerisation in the language operations of organisations that do not call for such far-reaching changes in procedures as does MT, but which none the less can improve the flow of work, bring about at least modest economies and at the same time serve as a

Application	Year 1	Year 2	Year 3	Year 4	Year 5
Referencing	XXXXXXXXXXXXXXXXXXXX	000000000000	-----		
Terminology	XXXXXXXXXXXXXXXXXXXX	0000000000	-----		
OA (WP + E-mail)		XXXXXXXXXXXX	0000000000	-----	
Telecommuting			XXXXXXXXXXXX	0000000000	-----
Input controls	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXXXX	XXXXXXXXXXXX	0000000000000000	
Output adjustments	XXXXX0000000000	-----			
Machine translation				XXXXXXXXXXXX	00000000

Key:	XXXXXXXXXXXX	Planning and design/training/first trials
	000000000000	Limited operation
	-----	Normal operation
	OA	Office automation
	WP	Word processing
	E-mail	Electronic mail

Figure 1. Timetable for the computerisation of language operations in a large organisation

training ground for staff who are not yet computer literate. Therefore, the best approach to MT in many organisations will be a gradual, 'layered' introduction of new technologies. By the time the subsystem interfaces on both the input and output sides have been adapted, the language services will have been able to digest the successive steps in progressive computerisation of their operations, which might range from the provision of referencing and terminology resources online, through word processing and electronic mail to telecommuting,* each of which can contribute towards the effectiveness, first of traditional translation operations, and ultimately of machine translation in the day-to-day language operations of the organisation.

If the elephant is thus eaten, one bite at a time, it will be certain to prove more palatable than if it were taken as one mammoth mouthful.

AUTHOR

Jean Datta, Reviser, English Translation Unit, and Officer-in-Charge, Reference and Terminology Unit, UNIDO, Vienna, Austria.

*Telecommuting: home-working, using a modem and telephone link to transfer work between a microcomputer in the worker's home and a computer on the employer's premises.