Another approach to streamlining complex language operations

by Jean Datta

Recently, the debate concerning improvement of translator efficiency through the use of machine translation has become quite heated. the proponents say it's revolutionary; the opponents say it's rubbish. What everybody agrees on is that you cannot just plug in MT, push the button, and sit back, watching all your problems get solved automatically. A great deal of dictionary building is required, just to begin with. And then, depending on the type of text and the purpose for which it is to be used, you have to lay on at least a certain amount of post-editing. Then you hope that the amount of translator effort you save will more than offset the effort you have invested – and are going on investing – in MT. In some cases, it may, in others, it may well not.

In view of the risk involved in installing MT in language operations of any scope and complexity, it behoves those in charge of such operations to endeavour to streamline their operations as much as possible before they contemplate introduction of MT. How such a streamlining can be achieved will depend entirely on the specific situation. However, the types of problems that might be targeted in this context can be illustrated by looking at the world-wide conference-servicing operations of the United Nations.

The United Nations itself – without even considering its specialised agencies – has offices in all parts of the world: New York; Geneva, Switzerland; Vienna, Austria; Santiago de Chile; Addis Ababa, Ethiopia; Nairobi, Kenya; Bangkok, Thailand, to name the most important. Some of these offices

organise heavy programmes of meetings for which extensive language services are required, both before and during the meeting periods. The workload must be forecast and short-term staff recruited.

This creates a dynamic and complex environment in which there can be "many a slip twixt the cup and the lip". Some of the obvious problems are:
- Over-recruitment: A conference may generate fewer pages of text than forecast, and staff may be sitting idle;
- Under-recruitment: A conference may generate more text than forecast after it is too late to recruit more people, so that the load is unmanageable;
- Time constraints: Urgent work may be generated late at night for completion early the next morning, and if no "graveyard shift" has been foreseen, it will be difficult to handle the work; also, provision of shifts around the clock means recruiting more staff, thus increasing staff costs considerably;
- Location constraints: Transporting people from Paris to New York, from London to Addis Ababa, and so forth, and paying for their hotel rooms and meals is very costly; these expenses multiply the cost of each page of finished text by a substantial factor.

One way to alleviate these problems would be to move the workload around instead of moving people around and laying on more people in locations where more work was expected to be generated, but might not actually materialise. Workload equalisation by means of the establishment of a world-wide

conference-servicing network with a node at each of the existing duty-station locations would make it possible to operate with a relatively small permanent staff and an absolute minimum of short-term peak-load help in each location. By sending texts for translation over a telecommunications link from a word-processing system in one location to a compatible word-processing system in another location, it would be possible to take advantage of variations in workload; work would move from locations where the load was heavy to locations where it was momentarily light, be done there and sent back to the origin for reproduction, distribution, etc.

Time-zone differences could also be put to use in such a scheme so that, for example, a job generated at 6 pm in Vienna or Geneva and needed for 8 am the next day could be sent to New York, where it would be noon, and done there during normal working hours. This would help to reduce the need for shifts and thus would make it possible to recruit fewer people on a short-term basis to cover contingencies arising out of conferences. In addition, skills in scarce supply could be shared by different locations, especially if no one location needed a given rare skill on a full-time basis.

The technology is available to realise such a scheme, in any case for languages using the Roman alphabet. The cost of transferring texts electronically between locations can be calculated. The savings could be estimated and determined more precisely through a pilot project of limited scope. However, as attractive as such a scheme may appear, it should not be foregotten that,

however the cost and savings calculations may turn out, existing work procedures would be profoundly affected by introduction of the scheme unless appropriate redesign arrangements were made in advance.

The main areas where redesign of procedures might be required relate to specialised terminology and provision of background information and documentation.

Why these areas are likely to need special attention becomes immediately clear if one considers, for example, that translators at the United Nations Environment Programme (UNEP), in Nairobi, have developed expertise chiefly in the area of environmental quality, and specialised dictionary and documents collections have been amassed to assist translators in dealing with texts relating to the subject areas of interest to UNEP. Let us assume that texts are transmitted electronically to Nairobi from Geneva, where the United Nations office accommodates, among others, the Economic Commission for Europe (ECE). Texts emanating from the ECE may deal with such matters as leading-edge coal-mining or steelmaking techniques, which would be quite "foreign" to the staff expertise and terminology/ documentation resources at UNEP. How would the translators determine the correct terminology to use? How would they find knowledgeable persons to consult on points of substance, with the originating office thousands of miles away? Questions such as these would have to be answered before this scheme could fulfil its promise. The solutions to these problems would not be simple, quick or cheap, but if the stakes were high enough, in terms of potential savings, it would be worthwhile envisaging them.

Since the translators could hardly be expected to develop expertise in every subject area that might present itself – from automotive engineering to trade law and the treatment of drug addicts – they would need to be provided with comprehensive and authoritative terminological resources, in the form of terminology data bases developed at the different network nodes that could be queried over a telephone line from any other node. To develop such data bases is a long-term undertaking that would entail not only substantial hardware and software components, but also and above all an extensive and an on-going human contribution.

At least one person in the existing reference and terminology structure at each location would have to be assigned to the job of collecting and screening terminology for entry in the data base, and that person would have to establish a more or less formal network of translators in the location to contribute terminology and determine the approved equivalents in the various languages.

In addition, there would have to be a feedback mechanism to ensure that errors or gaps in coverage were rectified, and this mechanism would have to extend to all the locations at which the term bases were being used. The most convenient means of establishing both the local screening network and the world-wide feedback network would probably be an electronic mail system (EMS).

Each local network member would have an electronic mail box in which the centralising terminologist responsible for screening at the location would deposit proposed terms. The local members would send their comments on proposed terms and also themselves contribute terms to the centralising terminologist's electronic mail box. World-wide, any translator at any location could pass on feedback to the centralising terminologist at that location, who would transmit it to his or her counterpart in the home location of the term base concerned.

EMS links might also be used to enable staff at one location to transmit queries of substance to submitting officers at other locations, but this would be rather unwieldy in comparison with the direct personal or telephone contacts possible within a single location, and would presuppose the installation of a very extensive system of communicating terminals at all locations, which the officers to whom they were issued would also have to be trained to use.

As far as background documentation is concerned, when one considers the millions of pages of documents that have been produced by the United Nations since its inception, one realises the staggering magnitude of this problem.

There can be no question of sending full sets of all documents in all languages to all locations. Therefore, it would be necessary to analyse the nature of the workflow subject to transfer between nodes and to establish limitative guidelines for the types of documentation which could be fed into the world-wide network.

It might be decided, for example, that certain classes of highly technical documents would not be fed into the network at all, and that emphasis in network operations would be on texts dealing with conference generalities. Once it had been decided what classes of texts would be fed into the network, it would have to be decided what existing documentation and going back how far would have to be provided to the various locations, depending on what types of work they would be sent, and in what languages. The documents determined to be necessary to provide background at the various locations would have to be reproduced in microfiche form, and sufficient numbers of microfiche reader/printer units would have to be installed at each location.

Current documentation, which would be typed on word processing systems and thus available in machine readable form, could be directly transferred to microfiche by means of a computer-output-microfiche units, rather than being taken off the hard copies. Thus, the documentation link in translation could be made paper-free, also helping to solve the problem of storage space for tons of documents.

The concept of a world-wide conference servicing network is technically sound and undoubtedly very promising in the context of large-scale geographically disperse conference operations.

However, it is clear that, just as in the case of introduction of MT, a great deal of planning and work procedure redesign would be required to ensure the success of such a scheme. In many areas of professional activity, computers open up alluring vistas for saving time, effort and money, but they can hardly ever be simply grafted on to existing work procedures.

If computers are going to usher in the millenium in language activities, then it will only be on the basis of the good old fashioned hard work required to lay the foundations. In this, as in so many other areas, computers offer a path to improved efficiency, but not a short cut.

Jean Datta is a reviser in the English Translation Unit for the United Nations Industrial Development Organisation in Vienna.