

MT happening at Georgetown university

You don't really have to know everything there is to know about augmented transition networks in order to be up-to-date on machine translation (MT). But it does help a lot. This was only one small message of the many conveyed by the Institute on Machine Translation which took place from August 3 to 9 at Georgetown University, in Washington, D.C., under the auspices of the Division of Interpretation and Translation of its School of Languages and Linguistics.

The Institute was ably led by Muriel Vasconcellos, director of the Machine translation program at the Pan American Health Organisation (PAHO), a regional office of the World Health Organisation, and by Veronica Lawson, a translator and consultant on machine translation who has carried out numerous assignments for the European Communities.

It had a workshop format, combining presentations by Dr Vasconcellos, Mrs Lawson and experts on individual topics with live demonstrations of MT systems and wide-ranging discussions with the presenters and among the participants. This informal format proved particularly fruitful in view of the varied backgrounds of the participants, who represented the fields of theoretical and practical linguistics as well as computer science, and came from international organisations, United States government agencies, academic institutions and private business.

The live demonstrations, of the Logos system for the German-English translation pair and of the Spanam/Engspan system in operational use at the Pan American Health Organisation for the Spanish-English and English-Spanish translation pairs, took place at the Wang Laboratories offices in Bethesda, Maryland, where excellent technical and physical facilities were made available.

Participants had the opportunity to submit samples of Spanish and English text to the Spanam/Engspan system and to appreciate how well it dealt with them despite the fact that it has been designed with a view to the narrow subject area of medicine and public health. A great deal of comment was attracted by the dictionary-updating feature offered by Logos, permitting unknown words found in a text during a preliminary scan to be entered in the system's electronic dictionaries by the human translator responding to interactive prompts before the translation is run.

The presentations covered the whole range of MT, from its history and a survey of operational MT systems, through the place (if indeed any) of artificial intelligence in the development of MT systems. One of the highlights was Dr Vasconcellos's presentation on "quick-fix" post editing, in which she revealed a few of the tricks of the post-editing trade and explained how she had programmed glossary keys on her word processor to provide ingenious instant solutions for often-required manipulations of raw MT output.

Another highlight was the revelation of the secrets of augmented transition networks by one of the PAHO computational linguists. For the uninitiated, the augmented transition network is the nifty mechanism which the PAHO MT algorithm uses to parse sentences,

and, if all goes well, clear up the part-of-speech ambiguities that are so common in English. In the sentence: "Sample figures show that slow sand filters produce more even results", for example, each of the words taken alone could be more than one part of speech, and yet using the augmented transition network, as presented at the Institute on Machine Translation, the *tour de force* of automatically analysing it becomes possible.

A third highlight of the Institute was the presentation on microcomputers and translation, which painted a picture of a general trend towards distributed processing, putting microcomputers on more and more desks, and evaluated the possibility of transferring existing MT systems operating on mainframes and minicomputers to microcomputers. It also sketched recent developments in such areas as firmware and chips specially designed for character-oriented tasks such as text matching and voice recognition, local area networks and the potential for the use of MT in dial-up information and other services.

Every topic covered by the Institute embodied a message for participants interested in ways computers can be beneficially put to use in the translation process. For example, the discussion of automatic sentence parsing and the coding of entries in electronic dictionaries led inescapably to the conclusion that machine translation can best be

applied where it is possible to ensure compliance with guidelines regarding the drafting of input text in relatively short, clearly structured sentences comprised of words used in their correct dictionary meanings. Unfortunately, these basic rules of good drafting are all too often disregarded, to the dismay of translators and end users of texts, alike. If MT could awaken a new concern for discipline in the use of language, that in itself would be an important contribution.

Another clear message that emerged in the Institute related to the evaluation of MT output. If machine-translated text is looked at with an eye for perfect elegance, then the point is being missed. An MT version is intended to be quick, cheap and "dirty", in other words, rough. The true measure of its value lies in its utility to the submitter. If a machine translation is delivered rapidly at low cost and serves the purposes of the submitter, it is a success, unless of course the satisfied submitter has been misled by errors in the translation and believes that it meets his needs better than it really does. This latter point was the subject of animated debate among the presenters and participants. It was pointed out that when the computer goes astray it usually does so in a blatant way that leaves the reader in no doubt, something that cannot be said about some of the plausible renderings of mediocre or inexperienced human translators. None the less, there was consensus that every MT text should either be conscientiously post-edited by a competent professional translator/post-editor or else should be accompanied by a warning to the reader that the text might contain substantive errors.

A third message deserving of mention that emerged in the Institute concerned the goals of an MT operation. The discussion of various applications of MT, for example by the United States Government at Wright-Patterson Air Force Base in Dayton, Ohio, and by PAHO itself, suggested that MT thrives best where goals are clearly defined and limited. Under the US Air Force programme, only Russian-English translation is being done, and only technical and scientific texts are dealt with. Apart from the limitation in text type, Russian is a highly inflected language and thus lends itself more than some languages to automatic parsing, and Soviet texts of the type concerned are usually well edited. Furthermore, the Air Force sets its sights firmly on providing rapid information to

Government analysts without concern for elegance of expression.

PAHO, too, has had limited goals. It has tackled only the Spanish-English and English-Spanish translation pairs, and has focussed on medical and public health texts. It has tried to balance the need for quality against the need for speed and economy by giving the end user an end product which is adequate to his needs.

The Institute conveyed a number of practical messages, but it did not shy away from raising a few loftier issues. The panel discussion on the role of artificial intelligence in machine translation advanced the concept of feeding world knowledge into the computer to enable it to translate with more finesse, on the basis of an "understanding" of the subject. Can the computer "understand" anything? Is there a single, unified body of world knowledge, a single reality? If so, would it be feasible to feed this whole body of knowledge into a computer system? The human brain is a parallel processor, and that is why we are able to tackle a problem from several angles at once, giving us a comprehensive approach that computers cannot – yet – achieve. Will the parallel processors of the fifth generation of computers, still operating on digital representations of information, in other words, strings of zeros and ones, be able to achieve this same feat? Will human translators soon become an extinct species?

Nobody at the Institute on Machine Translation seemed to be too concerned on this score. Perhaps the main message of the Institute was that people still have it all over machines, but that where the conditions are right we should welcome the very real assistance that the character-string crunchers can offer us.

The Institute was a "happening" in the best sense of the word, and one not to be missed by anyone with an active involvement in translation, since it offered pointers towards the future of the profession. In the final session of evaluation, the participants urged that, in view of the volatility of the machine translation field, the Institute should be repeated every two years or so. It is to be hoped that the organisers will take this good advice to heart.

Jean Datta, Reviser English Translation Unit United Nations Industrial Development Organisation Vienna, Austria