

Report

Japanese and European experts on MT

Over 250 machine translation (MT) experts from all over the world gathered in Munich in August, during some of the hottest days of the summer, to exchange views and experiences. The strongest Japanese contingent ever seen at an MT conference in the West made this one of the most important gatherings yet held in this field.

In his thought-provoking opening address the conference chairman, Christian Rohrer, of the University of Stuttgart, put the machine translation scene into context. In western Europe alone, he said, more than 150 million pages were translated last year. There were not enough translators, the costs were enormous, and MT was the only solution.

He also ruled out the use of one language, such as English, taking over as the vehicle of world communication, "because we must preserve the linguistic and cultural identity of our partners: we don't want to reduce our national languages to the role of dialects, limited to folkloristic purposes."

So many Japanese developers were interested in presenting their sys-

tems at the conference, he said, that there were not enough time slots for all of them, although they were represented in the accompanying exhibition. In comparison with Japan, Europe was investing much less in MT. Even with the funding of one major project (Eurotra) by the European Commission the overall amount spent in Europe was far below the Japanese figures. Mr Rohrer called for more co-operation, particularly in Europe:

Research on MT is very expensive. Therefore we have to find ways to share the research and development costs in the precompetitive phase. To give one example, every MT system which translates from German into another language will need an electronic dictionary of German. Why don't we pool our resources and develop a common German machine-readable dictionary? Or, if we think in terms of Europe '92, why don't we share the costs of electronic dictionaries for the languages of the European Community.

He also criticised the educational system in Europe, where the study of languages was still seen as part of the humanities, with a consequent alienation from computers; where computers were used, they were seen to have a purely instrumental function, and not to be a tool for creative research.

The more a user of an MT system knows about language, computers and computational linguistics, the more he can profit by the system. Most potential users today have an unrealistic picture of what a computer can do with natural language.

Mr Rohrer also pointed out the close relationship between MT and technical writing, and called for courses on technical writing to be established at university level, and for more research on technical language and sub-languages in general.

How can one produce texts which do not contain ambiguities? Can we make MT fully automatic by controlling the input and still

produce texts which don't bore the people for whom they are written? For me these are fascinating questions. Why shouldn't they also fascinate some students and professors of German, especially if the student finds a job afterwards?"

The keynote address was given by Professor Makoto Nagao, of Kyoto University, who had organised the first MT Summit in Hakone, Japan, in 1987, and who looked at progress since that time. He threw out a number of figures, including an estimate of the cost of translating a 125-word page, which came out at 2,500 yen for rough translation, 5,000 yen for high quality, 4,000 yen for average quality, 10,000 yen for Japanese to English high quality with specialised knowledge. He also estimated at 800 billion yen per year the current size of the Japanese translation market.

He set out seven tasks for the conference, to identify the state of the art, to promote MT, to examine how MT could be more closely matched to users' needs, to identify faults and means of overcoming them, to study natural language in order to secure better translation, to make a wide-ranging study of natural language processing in preparation for the information society of the 21st century, and to promote international cooperation.

He called for consideration of the establishment of an "International Association of Translation Technology", or an "International Association for the Promotion of Machine Translation", a call echoed by other speakers, including Loll Rolling, of the Commission of the European Communities, and by Christian Rohrer in his closing remarks, who suggested such an association could be created at the next conference, MT Summit III, which, it was agreed, would be held in North

America in 1991.

Individual systems or research projects which were the subject of papers at the conference were METAL (Siemens), Systran (Gachot), Logos (Logos Corporation), ATLAS (Fujitsu), HICATS (Hitachi), TAURAS (Toshiba), LMT (IBM), Rosetta (Philips), MU-2 (Japan Information Center of Science and Technology), JFY-IV (Academy of Social Sciences, Beijing), ODA (Tokyo Institute of Technology), CMU-MT (Carnegie-Mellon University, Pittsburgh) and Eurotra (Commission of the European Communities).

The audience learned that METAL was now being used at developed at 15 installations, in Denmark (Kolding), Belgium (Leuven), Germany (Munich) and Spain (Barcelona).

Replying to questions its presenter, *Thomas Schneider*, of Siemens, said that production gains, when the system was installed, dropped below a factor of one to one as users learned how to use it, but then rose again and went up to a factor of two to one.

Denis Gachot said that Systran now had 24 language pairs, of which 12 were operational, and they were now therefore emphasising its modularity. He also emphasised the importance of close relationship between user and developer.

Mr Gachot was questioned about the parameters for the quality levels which his company attributes to the translation in various language pairs at different stages in development. He said that the percentage was based on a qualification of 100 points, based in turn on the analysis of a corpus database of about 8000 sentences.

Bernard Scott said that the complexity of its rule base was a distinguishing feature of the Logos system. He stressed the importance of attributing generic categories when entering terms, so that a concrete noun would be identified as to whether it was a fabricated substance, and whether it was a

support surface — chair, desk, ledge, floor, shelf etc. — or a container, barrier, conduit, structural member, fastener, etc.

Logos was now installed in over 40 systems in Europe and North America.

Importance of generic categories

Fujitsu's ATLAS system, we learned, now had more than 100 users for its English to Japanese and Japanese to English language pairs. It used an interlingual approach. The Japanese to English system was first marketed 1987, and they were now experimenting with other languages. So far they had analysed and generated text in Japanese, English, French, German, Chinese, Swahili and Inuit with no modifications being necessary to the software, which indicated that the language-independent approach to morphological analysis in ATLAS gave it a structure suited to multilingual translation.

The HICATS system for translation being developed from Japanese to English incorporates 5000 grammatical rules, a basic term dictionary of 50,000 words, a technical dictionary of 250,000 words, and a user dictionary. Experience had taught the developers that if the source text was pre-edited drastic improvements in translation quality and cost-effectiveness could be obtained, and they were pursuing their examination of controlled language.

Toshiba was developing TAURAS as their experimental model, while their operational system was AS-TRANSAC. The developers placed great emphasis on the automation of the whole process, from inputting of the source text to outputting of the target text, and therefore regarded such items as Optical Character Recognition, Spelling Checks, Pre-editing and Desktop Publishing as all part of the system.

Michael McCord, who presented

IBM's research project LMT, showed how analysis was based on the Slot Grammar System, developed in the 1970s, but now used in combination with phrase structure grammar techniques and top-down parsing. Replying to a question as to when the development might be turned into a saleable product by IBM he replied, to laughter, "We're not allowed to even speculate on that".

The presentation of the Japanese ODA project led some in the audience to see it as an Asian parallel to Eurotra, in that it aims to provide a system for translating between Chinese, Indonesian, Malay, Thai and Japanese, using an interlingua approach.

We learned that the analysis, dictionary, generation, input and output and translation support systems had been "tentatively" produced and integrated into a "tentative" total system, and verification tests made using a limited number of sample sentences, and this was the stage so far reached. There was a hint that the developers were having difficulties with the interlingua approach:

A multilingual machine translation has many difficult problems. Especially the 'interlingua' approach seems to be a difficult research theme, and we should do more research on the theme. We should not abandon the interlingua approach because of the difficulties. Through conducting our project, many problems will be clear and we will recognise what is the most important problem.

Jan Landsberger, of Philips, Eindhoven, who presented the latest information on the long-running Rosetta research project, speculated whether Rosetta could be described as an interlingual system (the answer — in some senses yes, in others no). At present they were working on Rosetta 3, an experimental system able to translate short sentences and phrases between Dutch and English and Dutch and Spanish. It is planned to complete English to Dutch and English to Spanish phase translators in 1989. There will be

a second phase of development of Rosetta 3, and then the team will start work on Rosetta 4, a prototype system for a real application.

The MU-2 system is a development by JICST of the four year Japanese government-supported Mu Project which ran from 1982 to 1986. Test operation is scheduled to start in the next few months, and it may be possible to commence practical operation in 1990.

Information on JFY-IV revealed that it had been under development, for translation into Chinese from English, French, German and Russian, for the past 13 years, and that work was no proceeding on translating from English into Chinese. An experimental model was also being established for translation from Esperanto to Chinese.

Sergei Perschke, head of the European Community Eurotra project, decided to deviate somewhat from his paper as printed in the proceedings to give a spirited political defence of Eurotra.

After describing its organisational situation, he described it as one of the most scrutinised R & D projects of the Community, despite its modest size, modest in comparison with some other projects. "Why do our decision makers spend so much time on such a small problem", he asked. It was the first really major project since the 1966 ALPAC report. It had stimulated expectations, imagination, and fears. It had excited the imagination. An effort had to be made to maintain some degree of realism, and convince people they could not have everything immediately. "I believe", he said, "that Eurotra has made a contribution to the renaissance of the machine translation not only in Europe but also in Japan."

Mr Perschke talked about "post-Eurotra", the need to improve the interlinguality of interface structure, and the need for a global strategic Community programme concentrating on reusability of lexica, grammars etc.

Despite the spirited "attack is the best defence" nature of his presentation, Mr Perschke was unable to avoid the inevitable critical question from the floor. The equivalent of \$25 million had already been spent, he was told, and there were no practical results, while Eurotra was blocking funding to other, possibly more useful, projects. "My conception", replied Mr Perschke, "is that the existence of Eurotra has increased or led to an overall increase of funding in the world for machine translation. The money for Eurotra is additional to national and industrial funding... Our aim is to stimulate activities. \$25 million over eight years is not a big sum, it's just one single industrial project."

In addition to the presentation of individual systems there was a panel discussion on the practical application of MT, a paper by *Professor Nagao* giving a Japanese view of the future of machine translation, and panel discussion on government views on MT, and on new directions for MT.

The discussion on practical application was notable for a "devil's advocate" contribution by *Gerhard Freibott*, of Krupp, Duisburg. He criticised the exaggerated claims made by many of those marketing MT systems, and pointed out that comparatively few systems had been installed in industry. He ended by calling for objective criteria to be developed for measuring the efficiency and limitations of MT systems.

Other points that emerged in other discussions including the revelation by *Shigeru Sato* that just before leaving Japan he had tried to discover how many customers of the 100 who had acquired the ATLAS systems were using them, and the result was one quarter; the impressive degree of cooperation in Japan described by Professor Nagao, and the close involvement of the Japanese government in machine translation projects; and the insight into the growth in information and documentation — apparently more

information is now published in the world every year than was produced cumulatively in all the centuries up to the First World War.

Another maverick point of view was expressed in the "New directions panel" by M. Kay of Xerox Parc, California, who called for more tools for the translator, going beyond word processing and on-line dictionaries; more empirical studies of translation to be made ("no attempt whatever has been made as an empirical matter to discover what it is that translators do"); ways of escaping from "the determinism of batch translation", and the need to separate the deterministic from the heuristic; better ways of dealing with alternatives; the possibility sometimes of preserving ambiguities in the target text instead of trying to resolve them; concluding with the controversial remark, "If we are going to make MT better we are going to need to make it more expensive and slower".

The concluding session agreed that the next conference should take place in 1991 on the North American continent.

One aspect not mentioned from the floor, though much discussed in the lobbies and corridors, was the importance of having simultaneous interpreting between Japanese and English on a future occasion. The language of the Munich conference had been uncompromisingly specified as English, but while some of the Japanese speakers, such as Professor Nagao, have a good command of English, the presentations of some of the others were given in an English which was simply incomprehensible.