TACnology Corner

Special Report: MT Summit II

by Muriel Vasconcellos

More than 250 MT makers and shakers, users and watchers, traveling from 23 countries around the world, converged on Munich last August 16-18 to turn the second MT Summit into a major event. MT Summit II, following the example set by its predecessor held two years earlier in Hakone, Japan, offered a smorgasbord of updates on operational MT systems, reports on systems still under development, panels addressing the issues, and exhibits where software could be tested hands-on. The lavish Old World elegance of Munich's Bayerischer Hof made this banquet of treats seem all the more delectable.

Conference Chairman Christian Rohrer of Stuttgart University set the stage in his opening address by emphasizing that the primary aim of Summit II was to promote practical MT systems. The dismantling of trade barriers in the move toward Europe 1992 is only part of the general trend toward a global economy that already involves the worldwide circulation of information. "A world society needs a free flow of information," Rohrer said, "and it is translation which makes this flow of information possible." In Western Europe alone more than 150 million pages were translated in 1988. MT can be expected to play an increasingly important role, he believes, in processing the ever-larger bulk of text to be translated. MT Summit II was about practical applications of MT, about improving the quality of the input texts that are MT fodder, about ongoing research in MT and the contribution it can make to future high-level knowledge and information processing, about ways of sharing the cost of research and development, and, in the broad sense, about "promoting translation in general and machine translation in particular."

The keynote address, by "MT Superguru" Makoto Nagao, reviewed the expectations of the first MT Summit and the developments that have taken place in the interim, particularly with regard to practical implementation. What has changed the most, Nagao felt, are viewpoints and evaluations of MT. Contributing in part to the change in attitudes has been the study done in Japan which culminated in *A Japanese View of Machine Translation in Light of the Considerations and Recommendations Reported by ALPAC, U.S.A.* (the "JALPAC Report," see *Chronicle* June 1989). The final version of this 197-page document, complete with annexes, was distributed at Summit II.

Citing the need for a greater exchange of ideas and experiences, Professor Nagao called for the establishment of an International Association of Translation Technology, or an International Association for Machine Translation.

There were three sessions on operational MT and three on systems under development. The sessions on operational MT featured six systems.

METAL, presented by Thomas Schneider, is marketed by Siemens AG, which has also developed the 2 million-entry TEAM data base. The German-English version of METAL is now installed at more than a dozen sites, Schneider said, and after an initial break-in period of several months the users are achieving "considerable gains in productivity and a decrease in turnaround time."

The report on Systran, presented jointly by Jean Gachot and his son Denis, focused on the changes that had been implemented since Systran was acquired by the Gachot Group in December 1985, including the development of: new software, nine new language combinations, customer-specific dictionaries, and networking capabilities. Systran now has 24 language combinations, of which 12 are fully operational.

The LOGOS system was presented by Bernard ("Bud") Scott, who gave an overview of the system's history starting from 1973, an explanation of the motivations behind the different turns in its development, and a general description of the translation model. LOGOS is now a general-purpose multilingual system installed at over 40 sites in North America and Europe with full production versions available in six language combinations.

The ATLAS system, described by Hiroshi Uchida of Fujitsu Laboratories, is distinguished for being an operational MT system that uses an interlingua based on a concept structure. It has analyzed and generated text in Japanese, English, French, German, Chinese, Swahili, and Innuit (Eskimo) without modification of the software. In practical applications, translation by ATLAS combined with postediting has been found to take 30% to 50% less time than full manual translation.

Pre-editing was the subject of Hiroyuki Kaji's report on experience with Hitachi's HICATS (Japanese-

English). A controlled study was conducted comparing the times required to process (1) HICATS direct translation followed by postediting, and (2) pre-editing followed by HICATS translation followed by postediting. The latter scenario proved to be the most cost-effective.

In tests using Toshiba's TAURAS, Shina-ya Amano reported that total throughput, including input and postediting, depended on effective human performance and could range from 8 to 27 minutes for texts of approximately the same length, which were manually translated in 20 to 50 minutes.

In the panel "Practical Experience in the Application of MT Systems,"

. . . total throughput, including input and postediting, depended on effective human performance . . .

chaired by Veronica Lawson, a number of operational MT sites were reported or commented on. In the public sector, the European Commission was represented by Loll Rolling and the Pan American Health Organization by myself. Shigeru Sato mentioned several installations of Fujitsu's ATLAS, including the Mazda Motor Corporation (maintenance manuals) and Inter Group Corporation (scientific and technical abstracts for the European Commission). Other panelists were Reiner Arntz and Gerhard Freibott, who addressed such issues as where and how MT systems can best be used and why the technology is not more widely implemented. Contributions from the floor, including references to successful applications of other MT systems, added to the liveliness of the discussion.

A second panel, "Governmental Views of MT," chaired by Jan-Michael Czermak, reviewed the investments and commitments made to date by different governments, with policy positions on MT in some cases. The participants were Khalid Hussain (Kuala Lumpur), Akiro Kubota (Japan), Mahmoud Sieny (Saudi Arabia), Antonio Zampolli (Italy), Cornelis Jansen van Rosendaal (Commission of the European Communities--CEC), and Giovanni Coray (Switzerland).

The second half of the conference program was devoted to seven MT systems still under development. IBM's LMT, presented by Michael McCord, is a transfer-based system the shell of which is being used at different sites for five language combinations: English-Danish, Danish-English, English- French, German-English, and English-Spanish. Jan Landsbergen gave an update on Philips' interlingual Rosetta, the outgrowth of an English-language database inquiry project which expanded into the development of full MT using Montague grammar. Carnegie Mellon's KBMT-89, reported by Sergei Nirenburg, is, as its name suggests, a knowledge-based interlingua system. One of its special features is its augmentor component which helps to eliminate ambiguities by adding semantic and pragmatic constraints, and, if necessary, triggers an interactive dialogue with the user. Sergei Perschke explained the philosophy and provided highlights of progress to date under the massive multilingual EUROTRA project, toward which efforts are now under way at 19 sites in the 12 CEC member countries.

The Japanese government's Mu-2, reported by Tatsuo Ashizaki, already has a large Japanese-English dictionary (with a target of 500,000 entries by 1990) and will soon be tackling regular translation of the Japan Information Center of Science and Technology (JICST) data base. Zhuo Liu reviewed the different phases of JFY, the English-Chinese system under development in the People's

. . . greater attention to empirical studies of translation so that computational linguists will have a better idea of what really goes on in translation . . .

Republic of China for the last 13 years, the current version of which (JFY-IV) is intended for market. Hozumi Tanaka gave an update on the multinational undertaking of Japan's Center of International Cooperation for Computerization (CICC), already in progress for two years, which envisages the translation of multiple languages--starting with Chinese, Indonesian, Malaysian, Thai, and Japanese--through an interlingua approach. There were 12 exhibits at the conference, and the participants were able to see demonstrations of eight MT systems: Hitachi's HICATS, LOGOS, Siemens' METAL, OKI's PENSEE, Sharp EJ, Toshiba's TAURAS, Saarbr_cken's SUSY, and TOVNA.

The closing session, chaired by Christian Rohrer, provided a panoramic view of the future under the

Continued on page 28 . . .

MT Summit . . . *continued from page 21*

heading "New Directions in MT." Masaru Tomita of Carnegie Mellon University started out on the practical side with a summary of expected, or wished for, improvements in input and output. While conventional MT relies mainly on files as input, advances in technology should make it increasingly possible to enlist other modes. Of course, the ultimate goal is voice input and speech output, or real-time interpretive MT--an area in which Carnegie Mellon is already doing some work. Hard copy as input is another goal, which will depend on smarter optical scanners than we now have. A low-end type of MT could be developed based on options selected from menus by a mouse. And the keyboard itself will take on importance as sophisticated processing and parsing give renewed impetus to interactive MT.

A detailed model incorporating interactive MT was presented by Christian Boitet (University of Grenoble), who predicted the advent of small, specialized "light" MT applications based on the use of only core linguistic knowledge, specific domain knowledge, and intrinsic semantics. His model, LIDIA, would divide the processing between an IBM mini or PS2, on the one hand, and microcomputers with powerful editing capabilties, linked through a network, on the other. At the input level it would have the capability of interactive dialogue with the user, and ultimately it would have both written and oral outputs. Pierre Isabelle, of the Canadian Workplace Automation Research Center, stated the case for full reversibility of language directions.

Looking at the theoretical side, Margaret King (ISSCO, University of Geneva), foresaw a fundamental shift in pardigm, with insistence on declarativity and monotonicity. In plain English, this means that linguists will be writing programs in a declarative language without worrying about the order in which operations are carried out. At the end, a note of caution was sounded by Martin Kay (Xerox Parc), who advocated greater attention to empirical studies of translation so that computational linguists will have a better idea of what really goes on in translation and develop tools that will be more useful for the end-user.

At the close, Professor Nagao announced that he will be formulating a detailed proposal for the International Association.

MT Summit III will be held in the U.S. in 1991. Arrangements will be handled by the Center for Machine Translation at Carnegie Mellon University in Pittsburgh, and the meeting site is likely to be in or near Washington, D.C. □