TACnology Corner

A VIEW OF JAPAN AND A JAPANESE VIEW OF TRANSLATION TECHNOLOGY

ATA was represented on the program of the International Forum on Translation Technology (Oiso, Japan, April 26-28, 1989) by Veronica Lawson, Alan Melby, and myself. For me it was the experience of a lifetime, and I'm certain that my colleagues agree.

No pains were spared by the organizers – Conference Chairman Makoto Nagao; our generous host, the Japan Electronic Industry Development Association (JEIDA); and the organizing secretariat, Inter Group Corporation – to see that everything was perfect, from sumptuous repasts and accommodations at the Oiso Prince Hotel down to the last detail of the conference.

A View of Japan

First a word about the setting. Oiso is a resort town half an hour south of Tokyo by train, nestled on the edge of a wide bay with Mt. Fuji in the background. The hotel is on the beachfront, and our rooms looked out toward the bay and the ocean beyond, while directly beneath we could count ten different swimming pools, the largest of which had whirlpool jets and stretched along parallel to the shore as far as the eye could see. Fuji-San, which spends much of its time in the mist, became gloriously visible on the last day of the meeting and remained so until we had a chance to travel to Hakone on Saturday and rejoice in its magnificence at close range.

The conference, referred to as IFTT'89, was attended by more than 400 participants. The official languages were Japanese and English, and the simultaneous interpretation was superb. Thanks to our competent and indefatigable colleagues in the booth, we were all able to participate fully. In addition to the three days of plenary sessions, there were many excellent exhibits, including at least seven operational MT systems, all of them chugging away in English-Japanese or the more difficult combination, Japanese-English (one became very conscious of the differences, challenge-wise, between "EJ" and "JE"!). What impressed us Westerners most was the strong commitment to cooperation and mutual support among the Japanese MT developers.

The Japanese View

The primary focus of the conference was the JEIDA-sponsored study, *A Japanese View of Machine Translation in Light of the Considerations and Recommendations Reported by ALPAC, U.S.A.*, a preliminary version of which was presented on the first day by Professor Hirosato Nomura and discussed afterwards in a three-hour panel ably chaired by Professor Nagao.

This document, informally known as the JALPAC study, begins by reviewing nine steps of reasoning originally advanced by the Automatic Language Processing Advisory Committee in its report published by the National Research Council in 1966. The 20-year-old report claimed, as its basic premise, that (1) the supply of translators is already sufficient, and (2) the translation market is small. It then went on to assert that (3) the quality of MT is poor, and (4) MT does not necessarily lower cost. For these reasons, ALPAC concluded that (5) MT is not needed, (6) the cost of MT research cannot be justified, and (7) the basic technology for MT is yet to be developed. Accordingly, the Committee recommended that future efforts be focused on (8) further research in computational linguistics, and (9) strategies for the improvement of human translation. The JEIDA-sponsored study addresses each of these points in light of the situation as it now stands two decades later, especially from the Japanese perspective. Highlights from the document presented at ITFF'89 are summarized below, with occasional quotes taken from various points in its text.

As the first step in their analysis of the ALPAC arguments, the JEIDA group examined the current market for translation in Japan. A survey of 1,498 Japanese firms produced 121 responses. These companies alone (6.5% of the total) reported that they generate 1,966,242 pages of translation a year, most of it EJ or JE. It is estimated that all told the real annual total in Japan is 240 million pages. Nearly half this volume consists of product manuals and catalogs (48%), followed in importance by patents and contracts. Translators are paid an average of ¥2550 per page of draft copy for EJ and ¥3,600 per page for JE, or US$19.50 and US$27.50, respectively. (A page is 400 Japanese characters or 125 English words.) It is calculated that the cost of all this translation comes to ¥800,000,000 per year, or US$61,068,700. The volume is expected to double in the next two years. With these statistics for Japan, and given the needs for "large-scale in-depth communication in an international society," the study concludes that the situation is today "radically different" from what it was in the U.S. in 1966. It can no longer be said that there are enough translators, that the cost of translation is not a serious factor, and that the market is small.

The study points to changes in the nature of the demand as well. An increasingly large proportion requires highly specialized knowledge, whereas translators tend to be generalists. And there is a wider variety of languages. From the point of view of Japan and other countries, it is no longer "efficient" to consider English the "sole language of science," as was the case in 1966. Scientific advances are being reported in other languages, and

investigators cannot be expected to become polyglots in order to scan for the information they need. Another of the ALPAC claims was that the documents being translated were not being widely read. Today translations can be made available on-line, and with the growth of data bases MT can be expected to become "the major technology supporting the highly advanced information society of the future."

Moving on to ALPAC's claims about the quality and effectiveness of MT, the study cites advances over the last 20 years in computational linguistic theories, natural-language processing technology, and the development of linguistic data. This progress has been possible by dramatic improvements in computer technology and, at least in the Western languages, the widespread use of word processing, which facilitates both text input and postediting. Postediting, which was once cumbersome, is now a viable option – and less of it is required as MT systems grow and improve. Thanks to all these factors, machine translation has come of age, with systems now being utilized in the United States, Europe, and Japan. This statement is substantiated with references to successful MT applications around the world. Summing up its line of reasoning, the study concludes that the ALPAC Report "was mistaken both in the prospects for an increase in the demand for translation and in the progress of translation technology."

With regard to the current situation in Japan, MT is generally looked upon as "the basic technology for future high-level information processing." In addition to the government's official Mu2 project, MT systems are being developed by Oki, Canon, Sanyo, CSK, Sharp, Toshiba, IBM Japan, Systran, NEC, Hitachi, Fujitsu, Bravice, Matsushita, Mitsubishi, Ricoh, KDD, and NTT. The report does not specify which of these are in practical use. To date 4,000 MT systems have been sold – although "many are said to have been returned to the seller and some are not used and are idle." Hence it is difficult to know the real extent of MT utilization. In addition to full MT systems, there are also "portable translation machines of venture corporations," and there is research under way at universities. Telephony interpretation is another area of development.

Three kinds of MT service are being provided in Japan: (1) rough translation, (2) full translation using MT, and (3) VAN (value-added network) service based on MT. The first is a "rental service" in which raw translation is delivered direct to clients, usually for information only. In the second and third cases MT is embedded in the translation process, either using a system on-site or else tapping into a network, which can be done by freelance translators. With VAN service, the cost, including updating and postediting, comes to only 82% of the cost of human translation in Japan.

Satisfaction with MT in Japan has been hard to measure, but so far the largest positive response from users focuses on MT's value for the unification of technical terms and special expressions. The final version of the study will contain the results from a survey of MT users. In the meantime, there appears to be a consensus that MT technology is viable for the translation of: simple sentences, relatively simple embedded sentences, complex and compound sentences, sentences for which unique translation is possible, multiple meanings when the ambiguity can be resolved with a minimal addition of clarifying information, and clear syntax in which dependency is assumed to the nearest phrase.

The study ends up by identifying the tasks remaining to be done. Most important in the immediate future will be the gathering and incorporation of feedback from users. It is also urged that large corpora of natural language be collected and studied.

In the longer term, work needs to continue on terminology data banks and MT dictionaries in specialized technical fields, for the development of which an intergovernmental organization is proposed. This subject is reintroduced as a final recommendation.

Tasks in the field of computational linguistics, also the subject of a final recommendation, include additional work on: sentence analysis, the consistent representation of semantic information, representation of expressions in deep structure and their transfer, and sentence generation. In addition, principles from pragmatics and speech act theory need to be incorporated into MT systems, and it is essential that further work be done on dialogue comprehension – this being especially important for the telephony projects. As inferencing increases, there will be need for more efficient algorithms and for parallel processing. One of the final recommendations calls for further study of the human translation process, based on corpora of translated texts, in order to improve MT systems in their transfer and generation components.

Emphasis is placed on the need for multilingual translation as an expansion of current Japanese MT technology.

In the area of human resources preparation, the study calls for technical writing to be built into the science curricula so that professionals will produce clear texts that do not present problems for translation. In addition, it is recommended that there be further studies of controlled language. And finally, attention is called to the need to introduce courses in computational linguistics and machine translation in Japanese universities.

On the practical level, it will be necessary to step up the use of communication lines and electronic files. It will also be important to continue development of the translator workstation with supporting environments for pre- and postediting.

The foregoing points and related issues were discussed throughout the remaining sessions of the conference, which included many examples of current experiences with MT utilization. We Westerners were especially interested in news about the different Japanese systems. (A summary, "Machine Translation and Information Processing of the Japanese Language," is published as the 25th Anniversary issue of *Information Science*.) The

See TACnology, *p. 16*

TACnology Corner

Continued from p. 15

question of training posteditors was raised, and Inter Group reported that their translation school offers a three-month program specifically for the preparation of MT posteditors, which currently has 30 enrollees.

The agenda also included a report on the Center of International Cooperation for Computerization (CICC), a project for the development of a large multilingual MT system in five Asian languages: Chinese, Indonesian, Malay, Thai, and Japanese. This undertaking, which has government sponsorship in each case, has been under way since 1987 and involves the active participation of researchers and engineers from all the five countries. In Japan the cooperating parties are the Electronic Dictionary Research Institute (EDR), Fujitsu, NEC, Hitachi, Sharp, Toshiba, Oki, and Mitsubishi.

The conference heard differing views on expectations for MT, and it concluded with a series of blueprints for the future.

Muriel Vasconcellos, Chair
Committee on Translation and Computers

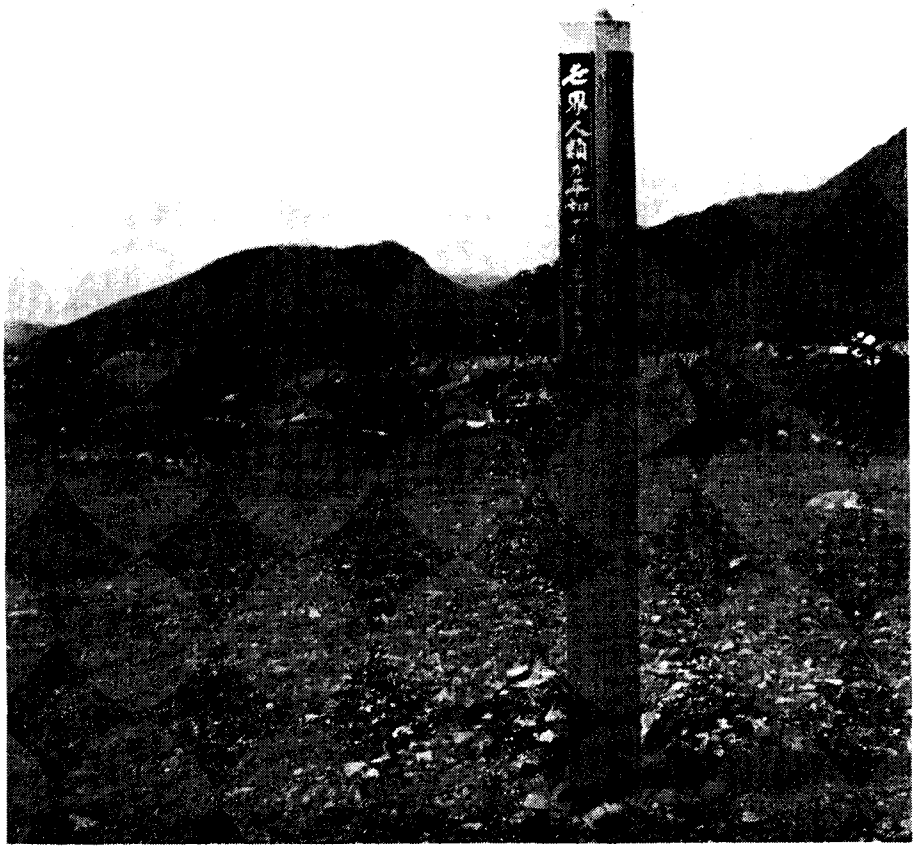

Photo by Muriel Vasconcellos

View of Mt. Fuji from atop Mt. Komagatake. Engraved on the stake is the message in English and Japanese: "Let peace prevail on Earth."