International Association of Machine Translation founded

MT Summit III, Washington DC, July 1-4

Is machine translation advancing fast, pushing ever forward into new areas of activity; or is it merely marking time; or is it even on the ebb? The place to find the answer to this question was obviously the MT Summit III Conference, held at the Mayflower Hotel, Washington DC on July 1-4. The MT Summit Conference, held every two years, brings together most of the world's leading MT developers, researchers and users.

The general impression left by this year's conference was that there had been no startling changes since the last MT Summit, held in Munich in 1989, but that there had been one or two incremental advances, particularly in Japan, and that more research, though most small-scale, was now underway. There were, however, one or two promising lines of development beginning to emerge.

Nearly 400 had registered for the conference, with a particularly strong participation from Japan. Japan was strongly represented also in the accompanying exhibition, where 22 developers were eager to demonstrate their systems.

The mood of the conference was decidedly upbeat. The chairman of one session, *Joseph Clark,* of the US Department of Commerce, called for a broadening of scope and perspective, and looked forward to future MT Summit conferences attracting participation in the thousands. In the same session, *Deanna Hammond,* president of the American Translators Association, confirmed that translation was now getting more attention in the United States than ever before.

Can MT fail?

One thing that was missing was anyone to act as devil's advocate. Indeed when Canadian translator *Claude Bedard* put in a plea from the floor for discussion of MT's failures as well as its successes, it was acidly firmly pointed out to him that if failures there were, which no-one was admitting, they probably resulted from misconceptions by the user.

Opening the conference, *Professor Jaime Carbonell,* of Carnegie Mellon University, said that the theme of the event this year was the user of MT. He listed the various new approaches to MT, such as neuro-networks. There was even work on machine learning. He looked forward to an extension of the scope of the conference to include OCR, document production, speech-to-speech, speech-to-text and text-to-speech research, and multilingual processing in using databases.

Timesaver

There was some discussion about the evaluation of the machine translation output, and one session was devoted to this, but as *Howard Teicher,* who chaired an earlier session on expanding the customer base for MT, remarked, the only practical criterion was the saving of the time of a translator using a system compared to a translator working unaided.

A representative of the Canadian government speaking from the floor, said that the volume of his translation work now amounted to 312 million words, of which 40% was contracted out to independent suppliers.

R. Fournier, of the Canadian translation vendor company, *Lexi-tech,* defended machine translation as a practical tool. For his company it was seen as a productivity enhancer — no more, no less. "It does work in spite of what people at these congresses say", he added.

Richard Kittredge, of the University of Montreal, talked about multilingual text generation, overleaping the need for translation, from raw data, giving weather information as an example. Text generation would probably be most useful, he said, in cases where the data was too volatile for human composition, for example, if the data was being updated every ten minutes. However, he suggested, the work done in this field could provide insights into the possibility of a deep, conceptual interlingua.

Statistical approach

The so-called statistical approach has been mentioned at previous conferences but in Washington it pushed itself to the forefront; probably this was the most striking aspect of the conference, and the strongest pointer to the future direction of MT.

Its foremost proponent was *Peter Brown,* of IBM. Speaking in the session "At the forefront of MT research", he presented his "probabilistic model" for translation. His research group was using the proceedings of the Canadian parliament, which

were produced in English and French. A statistical base of 98 million words of English and 85 million words of French had thrown up a database of 42,000 English words and 58,000 French words. No dictionary, no grammar, no linguistics were used in this approach, but out of 100 sentences translated in a test procedure, a correct translation was obtained in 50% of the cases, compared with a rate of 60% with the best commercial MT system. He was convinced that the statistical approach would take over in MT, since we were reaching the state where it would no longer be possible to type in all knowledge by hand.

Speech translation

The current state of speech translation was described by *Akira Kurematsu,* of ATR, Japan. A small vocabulary, of 400 to 1500 words, had been developed, for used with controlled (speaker-dependent, clearly pronounced) spoken language between Japanese and English. In five years the vocabulary should increase to around 3000 words of controlled spoken language in a specific domain. In ten years time he expected a vocabulary enlarged to something like 10,000 words, using spontaneous speech in limited domains.

Standardisation

Speaking in the same session, on work at the forefront of MT research, *Hozumi Tanaka,* of the Tokyo Institute of Technology, thought that development in multilingual machine translation would require the design of a better interlingua, and this would require more standardisation of language. Evaluation would be through multilingual corpora. During the next ten years he expected basic research would continue on semantic/contextual analysis, which was still the main problem in MT, and in using large dictionaries.

For *Masaru Tomita,* of Carnegie Mellon University, the great need was to educate users to make better use of machine translation.

Incremental advances (since similar papers at the 1989 MT Summit) appear to have been made at IBM, following on *Michael McCord's* LMT approach, with enhancements made in the Slot Grammar formalism, in the Taiwan-developed ArchTran English to Chinese system, and in the commercially available METAL and AS-Transac systems.

From Europe we heard about the moves towards EUROTRA-II, which is essentially a development environment seeking to capitalise on the research advances made in the European Commission's EUROTRA project, which involved the nine official languages of the Community. We learnt that the Commission is considering dividing the programme up into four phases (1) a mock-up phase (1991), (2) a development phase (1992-93), (3) a phase-in phase (1993-94), where the Commission would consider the possibility of producing a public domain version of the EUROTRA-II system, (4) and a shared cost phase (1994 and beyond), when it was hoped to move towards turning the system into a general purpose natural language processing and machine translation platform.

MT Association

The conference was used to launch a new International Association of Machine Translation, with *Professor Matoko Nagao,* of Kyoto University in Japan, as president, *Margaret King,* of ISSCO, Switzerland, as vice-president, *Dr Muriel Vasconcellos,* of the Pan-American Health Organisation, as secretary, *Vivian Merchant* (USA) as treasurer, and *W. John Hutchins* (UK) as editor-in-chief of a proposed newsletter. The new association will be a federation of three regional associations, the Japan Association for Machine Translation, formed recently; and the Association for Machine Translation in the Americas and the European Association for Machine Translation, which were both formed at separate meeting during the conference.