

Machine Translation (MT) SUMMIT VI - Past, Present, Future

San Diego,
California

29 October -
1 November
1997

If I had to think of a conference site more pleasant than San Diego I would be very hard pressed to do so. Unlike many conference venues, where the amenities may very well cater for all interests but you find yourself totally isolated from the rest of the world, the Catamaran Resort Hotel was situated just off a main thoroughfare riddled with interesting shops, bars and restaurants. The weather was every bit as magnificent as promised. Of course, the downside was that it took an extraordinary amount of self-discipline not to rush out of the conference hall to join the people that were surfing and roller-blading along the Pacific only yards away. This year marks the 50th birthday of Machine Translation, which, according to John Hutchins, MT's unofficial historian, saw the light of day on the March 4, 1947.

PRE-CONFERENCE ACTIVITIES

Four very useful pre-conference tutorials were offered, *A Gentle Introduction to MT*; *Theory and Current*

Vasconcellos, President of the International Association for Machine Translation (IAMT), formally opened the conference by cutting a ribbon at the entrance to the exhibition area. And what an exhibition it was! Twenty-six stands demonstrating the latest commercially available and prototype systems on the market.

DAY ONE

During the opening session, conference programme organiser, Scott Bennett, quoted John Hutchins by pointing out that "fifty years on, Machine Translation is alive, well, and used". Muriel Vasconcellos presented the newly established (IAMT) Award for Life-Long Contribution to Machine Translation to Prof Makoto Nagao of Kyoto University. In her presentation she said that Professor Nagao, the founder of the IAMT, embodies the spirit of the Association in every respect. Jane Zorilla and Chris Miller paid tribute to Deanna Hammond, who died in October. Scott Bennett said that the Summit theme of Past, Present and Future would be reflected in each day's plenary sessions with two tracks running most of the time, one concentrating on R&D and the other on more applied areas.

What followed next set the tone of the rest of the conference. Tom Pedtke's Keynote Address, *US Government Support and Use of Machine Translation: Current Status*, had exactly the same inspiring effect on the 300 participants as a coach's half-time pep-talk would have had on a football team. Mr Pedtke is Associate Chief Scientist at the US Air Force National Air Intelligence Center and his litany of ongoing government-sponsored MT initiatives clearly illustrated Bennett's statement, that MT "is alive, well, and used". We hope to publish a transcript of this keynote speech in next month's issue of *Language Today*.

John Hutchins' plenary session, *First Steps in Mechanical Translation*, recounted the early years of MT and was followed by one of the most unique panel discussions ever held, *The Pioneers of MT*. Victor Yngve, who had attended the very first MT conference, Igor



Laurie Gerber and Muriel Vasconcellos cutting the ribbon to the exhibition area to officially open MT Summit VI

Practice, Eduard Hovy, ISI; *Making MT Work for You*, Marjorie León, PAHO; *MT Evaluation: Old, New, and Recycled*, John White, Litton/PRC; and *Post-editing MT*, Karen Spalink, ITS. An excursion to Ensenada, Mexico and a visit to the San Diego Zoo Wild Animal Park were also available. On Wednesday evening a reception was held to celebrate the 50th Anniversary of Machine Translation. Laurie Gerber and Muriel



COREL Corporation exhibition Stand

Mel'chuk, Alexander Zhol'kovsky, Peter Toma, Chris Montgomery, Tony Brown, Roger Heller and Lol Rolling all took us down MT Memory Lane. Andrew Booth was unable to attend but did send a very entertaining video, which ended with him saying, "May you flourish with the funding of the various people that fund these things". Mel'chuk said that he supposed it was better to be called a "pioneer" than a "veteran". His tongue-in-cheek anecdotes included a rather bawdy, but hilarious, account of Vanya Mukhin, a one-time head of MT development in the former Soviet Union, who was, not surprisingly, also a KGB official. There came a time when it was felt that Mr Mukhin should really be 'Dr' Mukhin. So the necessary strings were pulled to publish his thesis, which, mysteriously, never saw the light of day. Dr Mukhin received his PhD and his thesis was immediately stored in a Top Secret archive with very limited access. At that time, the Soviet Union was famous for its lack of privacy. Courting couples had to resort to dark corners and the great outdoors, not really a viable option in the middle of a Russian winter. It just happened that a woman that worked in the vault where Mukhin's thesis was stored realised the privacy that Top Secret access could afford and often invited her boyfriend to share this privacy with her. The room totally lacked the furniture that courting couples might require so, employing the sort of resourcefulness that was required in those days, they decided to strategically arrange Mukhin's thesis on the floor; after all, no-one ever read it! Somehow, the arrangement was discovered and, when Mukhin was told about it, he simply replied, "What have your achievements done for humanity? Just look at mine!".

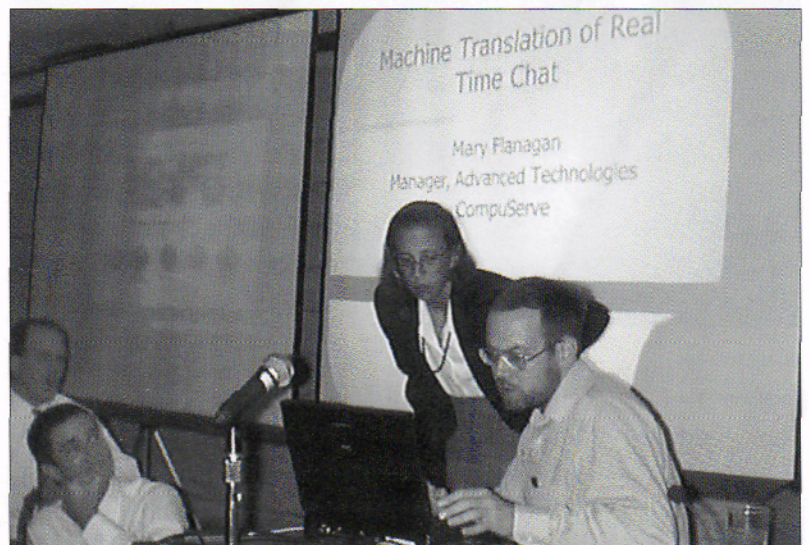
The parallel sessions that followed included *Machine Translation through Language Understanding*, Prof Makoto Nagao; *A Real-Time MT System for Translating Broadcast Captions*, Eric Nyberg & Teruko

Mitamura, CMU; *MT from the Production Perspective*, Mary Flanagan, CompuServe & Frederike Bruckert, LOGOS; *Managing Distributed MT Projects Today: A New Challenge*, Jennifer Brundage & Susan McCormick, SAP and Chris Pyne, S&D; *MT R&D in Europe*, Peter Quartier, Lotus & Roger Havenith, EC; *MT from an Everyday User's Point of View*, Annelise Bech, Lingtech; *Translating Scientific Texts Using MT and MAT Systems*, Vladimir Petrov, Polyglossum; and *R&D for Commercial MT*, Gregor Thurmair, GMS & Laurie Gerber, SYSTRAN. The day ended with a cruise around Mission Bay on the "Wm. D. Evans", the Catamaran's triple-deck sternwheeler and a Hawaiian luau on the beach.

DAY TWO

Harold Somers kicked off the second day of the Summit with his plenary session, *The Current State of Machine Translation*. Somers has the rare ability to be incisive, informative, and entertaining at the same time and this paper was no exception. He began by quoting a prediction that John Hutchins had made at the first MT Summit in 1987, "Predictions in the MT field have in the past been notorious for excessive optimism or for disillusioned pessimism. The safest prediction is that there will be a long-term future for nearly all varieties of systems. Machine aids, workstations, and interactive MT will be preferred by professional translators. Post-edited MT will continue in large translation services and in translation bureaux. Restricted input MT will remain the option for some multinational companies, unless 'raw' MT output improves greatly. Unrevised MT ... will be adopted increasingly for information-only translations. Automatic telephone translation seems unlikely in the near future, but universal access on

Mary Flanagan
presenting CompuServe's
Chat MT prototype



demand via public networks to many different kinds of automatic translation facilities ... does not now seem to be in the least utopian". Hutchins had been

accurate on all points apart from automatic telephone translation. As Somers points out, "His suggestion that universal access via public networks seems to have been the most prescient of all: in 1987 the Internet was largely an academic email network, and the size and scale of the World Wide Web was barely imagined at that time". Given that "predicting the future is always a pretty tricky business", Somers limited his paper to only 'predicting' the present. But, as he points out, "even that has become fairly difficult in the MT field. There was a time when it was realistic to expect to be able to attend more or less all the MT-dedicated conferences and workshops without access to a huge travel budget: they were spread all over the globe, but one could reckon on one, or at most two such meetings a year, at which one could expect to meet a subset of more or less the same 50 or 100 people. In addition, the literature on MT was fairly restricted, relevant articles, which appeared in journals dedicated to neighbouring or contributory fields like computational linguistics, translation studies, human-computer interfaces, software engineering, were infrequent enough that one could easily keep abreast of everything. And most significantly, if there were any commercial products to speak of, their appearance was heralded with a great clamour in MT circles. In short, with MT as one's chosen specialist subject, one could be fairly confident

software". And here, Somers echoed my thoughts exactly. Anyone that spends as much time as I do trying to keep abreast of translation technology developments will appreciate how much time, expense and effort it takes to mobilise as much technology and information under one roof as we found at this conference. Bearing this in mind, I would like to pose a question: Why is entrance to these exhibition halls always limited to conference participants? Great, over 300 people attended this year, a very admirable figure for a specialised event, but there must have been hundreds more potential users on the doorstep that may not have had the time or the inclination to attend the conference but would have dearly loved to have had the opportunity to visit the product exhibitions. Surely there must be a case for encouraging attendance by the general public either by ticket purchase or by vendor-distributed invitations. This approach would be particularly appropriate during the sessions, when exhibition attendance is virtually non-existent at some conferences. But, I digress. More pearls from Dr Somers: Other topics discussed in this paper were the role of linguistics in MT research contrasted with research in 'analogy-based' MT; automatic spoken language translation contrasted with text MT; tools, such as translation memory, bilingual concordances and software to check for mistranslations; Web use of MT; controlled language; and, an area of special interest to the speaker, the range of languages covered by MT, noting in particular the lack of development for minority languages. Somers recognises the obvious economic and political considerations when choosing a language for MT development: "Older readers will remember that researchers at Logos worked on Vietnamese and then Farsi before they settled on German. The influence of the funders is obvious here. Certainly within the European Union, less dominant languages such as Dutch, Danish and Portuguese have received attention, but it is still the 'big five' (French, German, Italian, Spanish, Japanese) that are targeted, and almost always paired with English. It is also interesting to note that the majority of systems are developed as bilingual systems, and although the development of a new language pairing might be 'piggy-backed' on the back of an existing system, truly multilingual systems are very much a rarity. There is still a big gap in the development of products (not just MT, but MT-related, and computer aids for translators) for minority languages, particularly non-indigenous minority languages". Somers is to be admired for drawing attention to what should be a painfully obvious vacuum in MT



Fujitsu Software Corporation Stand

of knowing pretty well what was going on all the time. This all contrasts with the situation nowadays. Commercial MT and translation-related products abound. We have a variety of annual and biannual meetings, including MT Summit, exhibitions that take much more than 10 minutes to visit, dedicated journals, newsletters and magazines, and a worldwide community of many thousands of people if we include - as we must - users (past and present) of MT

development and, hopefully, there might have been enough influential people listening to his paper to make a difference. Maybe even the folks that made up the panel discussion that immediately followed: *The Shape of Commercial MT Systems*. Each of the CEO's of LOGOS, SYSTRAN and Globalink had an opportunity to tell us where their companies are at the moment and where they hope to be going. Other parallel sessions included *Sharable Formats and Their Supporting Environments for Exchanging User Dictionaries among Different MT Systems as a Part of AAMT Activities*, Shin-ichiro Kamei, NEC; *JEIDA's Bilingual Corpus and Other Corpora for NLP Research in Japan*, Hitoshi Ishihara, JEIDA; *Multilingual Spoken Dialog Translation System Using Translator-Driven MT*, Hideki Mima, ATR; *User-Friendly MT: Alternate Translations based on Differing Beliefs*, David Farwell, NMSU; *MT R&D in Asia*, Hozumi Tanaka, Tokyo Institute of Technology & Keh-Yih Su, BDC; *PARS/U for Windows*, V. Petrov on behalf of Michael Blekhan, Ukraine; *From Metal to T1: Systems and Components for MT Applications*, Ulrike Schwall, GMS; *Production MT in the Public Agencies*, Dimitri Theologitis, EC & Joel Ross, FBIS. The day ended with the IAMT General Assembly followed by the conference banquet.

DAY THREE

Bente Maegaard's plenary session, *Whither MT?* completed the cycle of past, present and future. According to Dr Maegaard, "MT is going where the market and the users want it to go. MT will be available electronically over the network, and MT will be available in environments which also offer a variety of other tools for translation, as well as tools for other types of information management". Dr Maegaard agreed with Harold Somers: "Users of less spoken languages need MT and other tools just as much, or even more, than users of English or French. For some languages the market is not sufficiently large, which means that users of such a language will lack the tools which are otherwise available. This lack of tools will have an obvious economic effect, but also a cultural effect by excluding some languages from participating in an otherwise flourishing multilinguality". The final day's sessions included *MT R&D in North America*, Eduard Hovy, ISI & Elliott Macklovitch, University of Montreal; *Systran MT Dictionary Development*, Jin Yang, SYSTRAN; *MT as a Commercial Service: Three Case Studies*, Terence Lewis, Hook & Hatton; *Java and its Role in NLP & MT*, Tim Read, University of Granada; *End-to-end Evaluation in VERMOBIL I*, Rita Nuebel, IAI; *Using MT in a Corporate Setting*, Lou Cremers, Océ & Christine Kamprath, Caterpillar;

and *The MT Market Perspective*, Chris Miller, MCS, Colin Brace, LIM & Yukio Shimamoto-Caleo, Fidelity Management & Research. It seems to be a tradition for Sergei Nirenburg to fill the closing session slot with MT 'in' jokes.

In addition to the wide range of products being demonstrated in the exhibition hall, there were system presentations scheduled throughout the three-day period. The show-stopper was CompuServe's incredibly brave demonstration of their prototype Chat MT system. CompuServe's Chat facility allows virtual real-time text dialogue. Two computers went on-line, one with an English speaker, the other with a French speaker. Both computers were projected onto screens so that the audience could clearly see the results. Within seconds of one language being typed into its computer, the translation appeared on the other computer. This was all taken a step further by dictating the text into the computer. The translation appeared on the target computer between three and ten seconds later. People were invited to suggest phrases for translation and, miraculously, every suggestion was translated correctly. For the present, CompuServe is using Transcend as their MT engine.

This was probably one of the most optimistic MT events I have ever attended. Tom Peditke got everyone fired-up on the first day and the buzz seemed to last right to the end. But why was everyone feeling so optimistic? I think the major reason is that MT now seems to be accepted 'warts and all'. This is largely due to the dramatic growth of 'information only' users on the Internet. In development terms, there seems to be general agreement that systems should use 'everything that they can get their hands on', hybrid systems. In his paper, Harold Somers said that the problem of 'real-world knowledge' may never be overcome. I got the impression at this conference that the MT community has finally accepted this possibility and, like Sisyphus realising that he may never get that boulder to the top of the mountain, they seem to have placed their 'real-world-knowledge' boulder off to one side and are getting on with business. ■

Bob Clark



Dr Svetlana Sokolova,
developer of the Stylus
Russian MT system