

# MT News

## International

*Newsletter of the International Association for Machine Translation*

ISSN 0965-5476

Issue no.6, September 1993

---

### IN THIS ISSUE:

**Conference Reports**

**Association News (IAMT)**

**Users of Systems**

**Systems and Projects**

**Research Developments**

**Publications Received and Announced**

**Conference Announcements**

**Forthcoming Events**

**Application and Registration Forms**

**Notices**

---

### ***Editor-in-Chief:***

John Hutchins, The Library, University of East Anglia, Norwich NR4 7TJ, United Kingdom

Fax: +44 (603) 259490; Email: L101@uea.ac.uk

### ***Regional editors:***

**AMTA:** Joseph E.Pentheroudakis, Microsoft Corporation, Redmond, WA 98052, USA

Fax: +1 (206) 936-7329; Email: josephp@microsoft.com

**EAMT:** Tom C.Gerhardt, CRPCU/CRETA, 13 rue de Bragance, L-1255 Luxembourg.

Fax: +352 (44) 73 52; Email: tom@crpeculu.lu

**AAMT:** Professor Hirosato Nomura, Kyushu Institute of Technology, Iizuka, 820 Japan.

Fax: +81 (948) 29-7601; Email: nomura@dumbo.ai.kyutech.ac.jp

### ***Advertising Coordinator:***

Bill Fry, Association for Machine Translation in the Americas, 2101 Crystal Plaza Arcade, Suite 390, Arlington, VA 22202-4616, USA. Tel: +1 (703) 998-5708; Fax: +1 (703) 998-5709.

---

*Published in the Americas, with the generous assistance of Microsoft Corporation*

---

## CONFERENCE REPORTS

### MT Summit IV Features

### "International Cooperation for Global Communication"

*Muriel Vasconcellos and John Hutchins*

The Fourth Machine Translation Summit, held in Kobe, Japan, on 19-22 July 1993, lived up to its theme, "International Cooperation for Global Communication." The conference was attended by 220 participants from 15 countries. Both the keynote speech and a concluding panel addressed the theme specifically, while reports of several multinational projects highlighted the importance of international cooperation and attested to an increasing trend in this direction within the MT field. As Conference Chair Makoto Nagao pointed out in his opening speech, the International Association for Machine Translation has provided a framework within which such cooperation can be advanced and in fact accelerated. IAMT's impressive growth since its establishment two years ago at MT Summit

III is clear evidence not only of interest in MT throughout the world but also of the strong desire of researchers, developers, and users to cooperate at both national and international levels.

MT Summit IV marked a new beginning in more ways than one: it was the first Summit to be organized within the framework of the International Association for Machine Translation, and it was also the first to be held again in Japan, bringing the tradition of biennial MT Summits full circle and starting a new rotation between the three regions of the world. In addition, IAMT gained a new president at the close of the conference when the Association's founder and outgoing president, Makoto Nagao, turned over the gavel to Margaret King, president-elect of IAMT and president of the European Association for Machine Translation (EAMT).

The Hotel Okura Kobe site, facing the magnificent sweep of Osaka Bay, was a fitting venue for this major event. Indeed, one of the special treats of the conference was the welcoming reception, which took the form of a twilight cruise around the harbor.

In his keynote address, Makoto Nagao stressed the importance of MT being utilized appropriately. Citing surveys of MT use by the Asia-Pacific Association for Machine Translation (AAMT) and the Japan Electronic Industry Development Association (JEIDA), he pointed out that a sizable number of translation services count on MT to handle as much as 1,000 pages a month. The studies revealed that successful user sites tend to have the following characteristics: input is already in electronic form, the subject matter is highly focused, and thoroughly customized dictionaries have been built up over time. The break-even point for cost-effectiveness seems to come after about 10,000 pages a year. There continues to be a problem with the quality of both input and output, but the advantages of MT appear to outweigh these problems. Current emphasis is on greater utilization of networks, the development of filters between different word-processing and publishing systems, and the improvement of pre- and postediting facilities. He also noted the increasing use of MT on personal computers for information purposes only. By way of conclusion, Nagao outlined future tasks for all concerned with MT. Users should build dictionaries of specialized terminology, exchange experiences with others, collect typical corpora, cooperate in the development of guidelines for evaluation, and share all this information through user groups and clearinghouses. Developers, in turn, should clarify the information they give to the public, with specific statements about what their systems can and cannot do, and they should cooperate in the definition of common format codes for the exchange of texts and dictionaries as well as standardized evaluation procedures. And finally, researchers should explore new approaches such as statistical and example-based MT, cooperate with counterpart teams on a worldwide basis, and at the same time continue to press forward with discourse analysis, syntactic and semantic disambiguation, and the build-up of knowledge sources.

The conference featured three invited speeches. The first of these, by John Hutchins, covered the latest developments in MT technology. Since the end of the 1980s the most striking change has been the emergence of new approaches and methods in what is broadly called 'corpus-based' MT, particularly the increased use of statistical methods and explorations of example-based MT. However, the rule-based approaches characteristic of systems in the 1970s and 1980s have also witnessed considerable changes: the widespread adoption of constraint-based and unification formalisms, enabling the development of general-purpose shells for NLP and MT; the move towards 'lexicalist' approaches (illustrated at its most extreme by the 'shake-and-bake' approach), the increased attention to lexical acquisition, and the greater emphasis on the generation of idiomatic output. Other significant developments include multilingual generation from non-textual databases, research in systems for monolinguals not knowing target languages and numerous domain-specific systems developed by companies for specific purposes exploiting well-established methods and techniques of NLP and MT. He ended by speculating about the types of MT systems which may emerge from the various and sometimes divergent methodologies seen at present. He suggested that during the 1990s we may see the appearance of a "third generation" of systems, founded on a linguistic rule base (although less abstract than those typical of previous interlingua-transfer systems) incorporating a mixture of dictionary-derived lexical information, databases of domain-

specific knowledge, aligned corpora of bilingual texts giving examples of translations, and the use of probabilistic approaches to lexical and structural transfer and selection.

In an invited speech on the current state of machine translation usage, Muriel Vasconcellos reported data from a worldwide study of MT users. The survey yielded 40 responses, 23 of which gave annual production figures. The data from this subset alone showed that users are enlisting MT to produce more than 170 million words (680,000 pages) a year. On the basis of these figures and the number of other known users (about 80, most of them with smaller installations), Vasconcellos proposed that a very conservative estimate of MT use in the world would come to a total of 300 million words (1.2 million pages). Of the volume reported in the survey, 60% involved the translation of technical manuals, software, and other materials related to localization. At least one of these users found MT "indispensable," and those that reported on productivity cited increases ranging from 30% to 50%, which for them meant both reduced costs and faster turnaround. They were happiest with the fact that MT keeps terminology uniform and saves the need to re-enter format codes in each translation. The most frequent complaints had to do with document preparation, the quality of input texts, and postediting. [A revised version of this paper is reproduced in this issue]

The third invited speaker was Toshio Yokoi of the Japan Electronic Dictionary Research Institute. His topic was the problem of constructing very large-scale knowledge bases, not just for MT but for a multitude of natural language processing purposes. He began by stressing the crucial role of natural language for the representation of knowledge, not just as the interface between knowledge databases and their users. Major requirements of large knowledge bases must include: a) the capacity to expand without difficulty, to enhance the quality during expansion, and to absorb information from a variety of different sources; b) support for interpretation by both humans and by computers, for integrating representations in many types of languages (natural and formal), images, diagrams, sounds, etc.; c) demands on the development of new system architectures, where natural language is the core medium of the computer system; d) the ultimate goal of full understanding of natural language in reliable processing. The EDR electronic dictionary project was given as an example of progress towards these ends. In one respect, the electronic dictionary was itself a large knowledge base, and here Yokoi stressed the relationship between the EDR concept dictionaries and the text base providing example definitions. More important perhaps was the experience gained during the project in lexical knowledge acquisition, providing techniques and methods for building knowledge bases in general. He concluded by describing plans for the "Knowledge Archives" project, the creation of a very large database of 'knowledge documents' - in free or controlled natural language, in formal knowledge representation languages, in pictures, images and sound.

The EDR project was the theme of other contributions – by Seibi Chiba and Yoichi Takebayashi - who together brought participants up to date with its current status. Hiroshi Yasuhara (another EDR member) concentrated on a description of the research on example-based MT using text data gathered for the dictionary. Research and development on text corpora was the central topic of Yorick Wilks' presentation. He gave a broad survey of developments in message understanding (SRI, BBN, Massachusetts), tagging of corpora (Lancaster, AT&T, Pennsylvania), alignment of bilingual texts (Brown et al., Church, Kay), stochastic grammars, machine-readable dictionaries (LODCE, CoBuild, etc.), connectionist models, large text corpora (Lexical Databank Consortium, the European Corpus Initiative, etc.), ending with speculations about the role of corpus-based approaches in MT, which was dealt at greater length in his paper in the proceedings which concentrated on an assessment of the achievement of the IBM Candide project.

As in previous years, the MT Summit provided a forum for the description of major MT and NLP projects. Susumu Funaki described the status of the CICC multilingual project involving teams from Japan, China, Thailand, Malaysia and Indonesia; and Meying Zhu (in a joint paper with Hiroshi Uchida) gave a detailed account of the structure of the interlingua being used. Ahmad Zaki Abu Bakar described the Malaysian contribution to this project and also outlined other MT activity on the translation of textbooks and of news bulletins for the Kuala Lumpur Stock Exchange.

The ATR project was described by Tsuyoshi Morimoto and Akira Kurematsu, with particular emphasis on its experimental speech translation system. In earlier versions more than 90% of

utterances were recognized and translated accurately in about 25 seconds processing time. The latest version (ASURA) has been greatly extended in vocabulary and sentence types, but accuracy has dropped to about 60% and processing time has increased to 50 seconds. The speakers were confident, however, that hardware improvements will enable the achievement of near real-time processing. In brief, ATR is claimed to have demonstrated the technical feasibility of an "Interpreting Telephony System" in the near future.

The third major project is also focused on spoken language translation. This is the Verbmobil project described by Wolfgang Wahlster of the German Research Center for Artificial Intelligence. The aim of this long-term project is the development of a portable device for aiding the translation of spontaneous spoken language in face-to-face negotiation dialogues. The first version will aim to provide translations on demand for participants (German or Japanese) who both have passive knowledge of English but are not fluent speakers; i.e. where English is used as a common language in business or technical discussions. The project is planned for 8 to 10 years; in the first four years funding will amount to 60 million Deutschmarks. International collaboration is planned with ATR and with three US groups at Carnegie Mellon, Stanford (CSLI) and Berkeley (ICSI).

A more general view of European activity in MT and related spheres was provided by Loll Rolling (Commission of the European Communities). He described the various projects which have been supported by the Multilingual Action Plan, the Eurotra project, the Language Research and Engineering (LRE) project, the ESPRIT programme, EUREKA projects such as GENELEX and GRAAL, and the ECLAT programme in language technology.

A panel on the evaluation of MT gave rise to a lively discussion with considerable audience participation. Under the chairmanship of Margaret King, some of the key issues in this area were addressed by Jaime Carbonell, Hirosato Nomura, Loll Rolling, and Muriel Vasconcellos. The panelists agreed that there is no single "right" way to evaluate MT and that there should be a variety of methodologies, much like a cook's batterie de cuisine, from which several can be chosen in combination, depending on the circumstances. They went on to discuss the differences between progress, adequacy, and diagnostic evaluation, as well as the best approaches for each. A heated debate ensued over the relative merits of glass box vs. black box perspectives. The conclusion seemed to be that a view inside the glass box is essential for progress and diagnostic evaluations and that a "gray box" may be best for evaluations that focus on adequacy: even though users may be more concerned with black box considerations, it is always important to know why a system fails.

Future developments in MT technology were the subject of a panel chaired by Hiroshi Uchida. Six experts took on the challenge of this topic: Christian Boitet, Pierre Isabelle, Hwee Boon Low, Sergei Nirenburg, Christian Rohrer, and Junichi Tsujii. The general agreement was that future MT systems will be 'hybrids', combining the best features of rule-based approaches, whether linguistic rules or knowledge bases, with the more recent stochastic and example-based methods. There would be more attention to specific user needs in the design of actual systems (i.e. fewer general-purpose systems) - selecting the 'best' methods for the purpose. Isabelle argued that corpus-based approaches were more appropriate to machine-aided translation than rule-based methods; indeed, research should concentrate on tools for aiding translation rather than systems for producing translations. Rohrer saw a major role in the future for multilingual text generation from databases; and Boitet speculated about the future for speech translation, the place of MT in multimedia communication (pointing to experiments in TV subtitling as an example), and aids for simultaneous interpretation.

In keeping with the theme of the conference, the third panel brought the substantive program to a close with an overview of current international cooperation in MT and plans for the future. Y.T.Chen, director of the NSF information center, gave the US view, emphasizing the advances in many areas related to language technologies and the growth of the global economy, multinational companies, telecommunication networks, and the appearance of new information- and knowledge-based products. He then outlined the support of the US government to research in these areas, through ARPA and NSF, and the encouragement of international cooperation, particularly the sharing of data and tools, the establishment of standards and evaluation methods, through demonstrations of

technology, workshops, joint sponsorships, and the establishment of the national information infrastructure. The Japanese view of was given by Makoto Nagao. He highlighted first the areas which are still problematic and where international cooperation is desirable, even essential: discourse, selection of texts suitable for translation, terminology problems, multilingual text corpus, learning mechanisms, parallel architectures for MT, methods of evaluation, networking MT systems, common text and dictionary formats for interchange. He ended by surveying current Japanese government involvement in the multilingual cooperative project (CICC), cooperative development of terminology dictionaries (EDR), and support for MT and NLP information exchange centres. A weakness in the Japan picture was the poor progress in building large text databases, with the main impediment being the problem of copyright. He thought also that far more discussion on standardisation was desirable in Asia; there was much activity here in Europe and the United States but Asian countries had lagged behind. Loll Rolling sketched previous CEC involvement in MT, and then went on to describe the present activities of EAGLES on cooperation with text corpora, lexical resources, evaluation and raising of awareness. He emphasized the need for projects and plans to be adequately reported and distributed, and believed that MTNI could and should play an important role in the dissemination of information about international activity. Yang Tianxing described the support for R&D in MT in the People's Republic of China, surveying briefly the current centres and their research activities, and describing China's involvement in the multilingual CICC project. Toshinori Saeki of MITI (Japan) stressed the large-scale support for CICC project by MITI, but he also pointed out the relatively low level of telecommunication networking in Japan and the need for its improvement if international cooperation is to be encouraged.

The final event was the Second General Assembly of the International Association for Machine Translation. The minutes of the Assembly are published in this issue. [Copies of the proceedings are available from the MT Summit Secretariat c/o AAMT, Akasaka Chuo Mansion 305, 7-2-17 Akasaka, Minato-ku, Tokyo 107 Japan. Tel: +81-3-3479-4396; Fax: +81-3-3479-4895.]

---

## **TMI-93 continues the 'rationalist-empiricist' debate**

*John Hutchins*

For three days (14-16 July) in the week preceding the MT Summit, researchers gathered in Kyoto for the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI '93). Its theme was announced as "MT in the Next Generation" and, being held just one year after the previous TMI conference in July 1992, it was perhaps inevitable that papers would take up the issues raised in Montreal, in particular the 'confrontation' between the well-established methods of rule-based approaches to MT system design and the newer corpus-based approaches (stochastic methods and example-based MT).

The first three sessions (on 14 July), were devoted to the newer methods. The opening session was started by Jing-Shin Chang and Keh-Yih Su with a description of the BehaviorTran (previously ArchTran) MT system at the National Tsing-Hua University, Taiwan ('A corpus-based statistics-oriented transfer and generation model for machine translation'). Its combination of rule-based and corpus-based methods illustrates an increasingly common option in MT system development. They were followed by Hiroshi Maruyama, Shiho Ogino and Masuru Hidano on 'The mega-word tagged-corpus project' for building a large, high-quality text corpus for Japanese. The main emphasis of the paper was a description of the Japanese morphological analyzer. Eiji Komatsu, Jin Cui and Hiroshi Yasuhara of the Electronic Dictionary Research Institute then reviewed the EDR project and its relevance to interlingua-based MT ('A mono-lingual corpus-based machine translation of the interlingua method').

Five papers on example-based MT (EBMT) followed. First, Sergei Nirenburg, Constantine Domashnev and Dean J.Grimes (Carnegie Mellon University) described 'Two approaches to matching in example-based machine translation' (comparing performance using different passage lengths); then, Satoshi Sato of the Japan Advanced Institute of Science and Technology spoke on 'Example-based translation of technical terms', describing the latest version (MBT3) of a series of

systems exploring 'memory-based translation' using an external database of translation examples, a Japanese thesaurus and an English thesaurus. The last talk in this session was by Stephen Richardson (in a paper written jointly with Lucy Vanderwende and William Dolan) on 'Combining dictionary-based and example-based methods for natural language analysis' which described experiments at Microsoft towards 'hybrid' NLP using rule-based methods for structural information and the extraction of information from dictionary definitions to create a large database of examples illustrating semantic relations. The final two papers of the day were devoted to aspects of preference scoring for selection in EBMT. Eiichiro Sumita, Osamu Furuse and Hitoshi Iida (of ATR) reported on 'An example-based disambiguation of prepositional phrase attachment'; and Kiyoshi Yamabana, Shin'ichiro Kamei and Kazunori Muraki (of the NEC Corporation) also spoke about resolving ambiguity problems in their paper 'On representation of preference scores'.

On the following day (July 15), it was the turn of the 'rational approach' in the first session. Jörg Schütz and Bärbel Ripplinger began with research emanating from the Eurotra project in a paper on integrating computational terminology in a constraint-based NLP/MT environment ('Machine translation supported by terminological information'). They were followed by Nili Mandelblit from the University of California, who argued for the building of interlingua MT on a cognitive linguistics basis by incorporating a knowledge base of language usage ('Machine translation: a cognitive linguistics approach'). The proposal was illustrated by translations from English into Hebrew. The third paper by Margherita Antona and Jun-ichi Tsujii put forward a rule-based approach to tense and aspect, involving a bilingual translation knowledge base ('Treatment of tense and aspect in translation from Italian to Greek - an example of treatment of implicit information in knowledge-based transfer MT').

There followed a session for papers describing other new approaches in rule-based paradigms: Koichi Takeda (IBM Research, Tokyo) on 'An object-oriented implementation of machine translation systems' (essentially a knowledge-based rule-driven approach within the research on Shalt-2); Hagyu Lee and Yung Taek Kim (Seoul National University) on 'An idiom-based approach to machine translation' (Korean-English); and Tetsuya Nasukawa (IBM Research, Tokyo) on 'Discourse constraint in computer manuals' (an examination of the sublanguage in this area, and using no extra-linguistic information).

The third session of the second day began with a description of 'Recent advances in JANUS: a speech translation system' presented by J. Tsutsumi on behalf of M. Woszczyzna and others from Carnegie Mellon University. It was this system which was involved in the demonstration of spoken MT by ATR in February [see elsewhere in this issue]. This was followed by Pierre Isabelle with a joint paper with others from CITI, Canada (previously CWARC) on 'Translation analysis and translation automation'. He argued that the analysis of previously translated texts can be analyzed as structured databases of great value as aids for human translators. He described TransSearch (a tool for bilingual concordances), the development of TransCheck - a system for detecting deceptive cognates ('false friends') - and research on TransTalk for facilitating speech-to-text transcription of dictated translations.

The final session of the day was devoted to a panel discussion on 'Translation equivalents' with contributions from Christian Rohrer, Jaime Carbonell, Yorick Wilks, Pierre Isabelle, and Makoto Nagao. The discussion was dominated to some extent by the theoretical question of whether 'equivalence' as such is possible, and reached the obvious conclusion that in MT a pragmatic approach must be adopted (thesaural, statistical, example-based, etc.) The discussion did raise some major issues: the inevitable finiteness of databases, the limits of compositionality, the abstract nature of dictionaries, the most appropriate unit of equivalence, the role of examples, and so forth.

The third and last day of the conference (16 July) began with three papers on source language analysis. Masaki Murata and Makoto Nagao (Kyoto University) discussed the 'Determination of referential property and number of nouns in Japanese sentences for machine translation into English'. Then, Satoshi Shirai, Satoru Ikehara and Tsukasa Kawaoka (NTT Laboratories) described experiments on a method for rewriting modal and tense expressions and expanding coordinate structures in Japanese input sentences, for implementation in the ALT/JE system ('Effects of

automatic rewriting of source language within a Japanese to English MT system'). Satoshi Kinoshita, Miwako Shimazu and Hideki Hirakawa (Toshiba Corporation) reported on methods for deriving reliable semantic data from texts, which can be applied in the analysis of ambiguities in other texts ('Better translation with knowledge extracted from source text').

The second session of the last day was devoted primarily to questions of MT evaluation. Masuru Tomita and colleagues from Keio University spoke about attempts to evaluate MT performance on translating English TOEFL (Test of English as a Foreign Language) texts into Japanese ('Evaluation of MT systems by TOEFL'). Evaluation by student examinees was not in general reliable, in part because the translations themselves were poor, but some guidance for devising evaluation tests was derived. This was followed by Adriane Rinsche who reflected on her experience in evaluating a number of MT systems on behalf of the Commission of the European Communities ('Towards a MT evaluation methodology'). She concluded that few truly quantitative measures are practicable and that some subjectivity is inevitable in this area. The third paper in the session was by Kwangseob Shim and Yung Taek Kim (Seoul National University) who described a statistical approach to structural disambiguation of verb phrases ('Towards a machine translation system with self-critiquing capability').

The final session of the conference considered problems of extracting information from corpora - thus returning neatly to the topics of the first two sessions on the opening day. Hideo Watanabe (IBM Research, Tokyo) described 'A method for extracting translation patterns from translation examples'. Shinichi Doi and Kazunori Muraki reported on NEC's statistical method for selecting translation equivalents ('Evaluation of DMAX criteria for selecting equivalent translation based on dual corpora statistics'). Finally, Teruko Mitamura, Eric Nyberg and Jaime Carbonell described the approach at Carnegie Mellon University towards the building of large knowledge bases for interlingua systems such as KANT and its successors ('Automated corpus analysis and the acquisition of large, multi-lingual knowledge bases for MT').

For those participants spending time in Kyoto before going on to the MT Summit in the following week, there was the fascinating glimpse of traditional Japan provided by the spectacular Gion Matsuri procession on the next morning. It was a most fitting climax to what was undoubtedly one of the best conferences in the TMI series. The next TMI conference will take place in 1995 in Louvain, Belgium.

Copies of the proceedings of TMI '93 are available from the organizers: TMI '93 Secretariat, HS Building, 7-1-9 Nishi-gotanda Shinagawa-ku Tokyo 141 Japan; or Hozumi Tanaka (Dept. of Computer Science, Tokyo Institute of Technology, 2-12-1 Ookayama Meguro Tokyo 152 Japan) and Yuji Matsumoto (Graduate School of Information Science, Advanced Institute of Science and Technology, Ikomashi, Nara 630-01 Japan).

---

## **Translation Fair '93**

*[From: AAMT Journal, no.2, February 1993]*

Translation Fair was held from December 15 to 17, 1992 at Science Museum in Kitanomaru Park in Tokyo, under the theme "Future Electronic Communication." The Fair was supported by various ministries and agencies, including the Ministries of International Trade and Industry (MITI), Foreign Affairs, Labour, Education, and the Science and Technology Agency.

The purpose of the Fair was to explore the future state of communication through discussions on various techniques, to find a way of effectively using machine translation systems, which are tools for electronic translation, and through the presentation of up-to-date information on research and development.

The Fair was co-ordinated by the Japan Translation Association (JTA), with special cooperation by the Centre of the International Cooperation for Computerization. These organizations were involved so as to encourage the participation of a diversity of groups, ranging from researchers and manufacturers to end users of MT systems. The Japan Translation Association is a corporate

body under the administration of the Ministry of Labour. The Association trains a large number of translators and gives translator certification tests as well.

The Center of the International Cooperation for Computerization (CICC), administered by MITI, offers assistance to developing countries in terms of computerization and promotion of various systems, as well as knowledge. Since 1987, under commission from the Japanese government, the organization has been jointly developing advanced multi-language machine translation systems with China, Thailand, Malaysia and Indonesia.

### **The opening ceremony**

For the opening ceremony on December 15, guests, the representatives of organizing corporations, and then persons concerned gathered at the venue. After speeches by the leading officials of the MITI and the Ministry of Labour, Mr. Nagao, the chief organizer of the Fair, Mr. Yuasa, chief executive committee member, and some distinguished guests cut the tape to declare the Fair open.

### **December 15: Symposium 1: Multi-Language Translation System**

Symposium 1 consisted of three sessions. First, Prof. Tanaka (Tokyo Institute of Technology) gave a keynote speech on "Research on machine translation and its future tasks to meet globalization." In this speech, prof. Tanaka introduced past and present trends in research and development of machine translation in various countries. He then revealed that CICC employs the interlingual approach instead of the transfer approach, which is currently on the market. The reason for this choice was: 1) the interlingual approach offers an easier means of developing multi-language translation methods; 2) various AI technique are integrated with this approach; 3) therefore, introducing this approach may have greater impact on technological development in developing countries. In fact, using this approach, these countries were able to construct particular systems for their own languages. Moreover, dictionaries made for machine translation are valuable not just for this original purpose but for understanding natural language processing. Important future tasks for researchers include the construction of large-scale knowledge bases and the integration of such techniques as human interfacing, evaluation method, standardization, networking, and parallel processing. The outline of the CICC projects was then introduced, followed by reports on the status of research and development in joint research countries: Thailand, China, Malaysia and Indonesia. Though these countries face different challenges in diverse environments, researchers in all of these nations have the same position and aspiration to achieve success in this field.

After Prof. Tanaka's speech, a panel discussion was held by the representatives of various institutes from these countries on the theme "International cooperation toward a multi-language translation system."

### **December 16: Seminars: Full Utilization of Machine Translation Systems for Users.**

#### *Session 1*

The theme of the first session was "The gradual introduction of machine translation systems and utilization know-how." In this session, Mr. Makino (associate professor at Toho Univ.) discussed improvement of machine translation process, technique and quality, from the viewpoint of engineering. Mr. Ogawa (Nikkei Printing Inc.) then said: "Those who think that machine translation can be used without editing know nothing about translation. At the same time, those who think machine translation works know nothing about MT." He then introduced the know-how involved in constructing dictionaries. Mr. Sakurai, representing the machine translation industry, said that even professional translators cannot do a perfect job, and explained the way of Japanese sentence edition. Mr. Mori (Babel Inc.) spoke on the important points in utilizing machine translation systems, from the viewpoint of training translators and editors.

#### *Session 2*

The theme for the second session was "This is how I use machine translation." In this session, Mr Harada (IBM) gave a detailed explanation, based on his five-year experience as a user, of the points



that need improvement in operability and design, and the improving process. He emphasized that to fully utilize machine translation systems, it is important to know the characteristics of the material to be translated and the particular flow of translation of that material, and to arrange the machine translation program, related programs, and operating method suitable for the material.

Mr.Ando (Toin Corporation) showed some examples of actual usage of machine translation, saying, "You have to use your brain just a little to use machine translation fully." Mr.Ozawa (NCR Japan, Ltd.) said that he uses machine translation to make a draft. In this way, he said, he tries to give an answer to the long-term conflict between competent human translators and machine systems. He also said, "If translators begin to employ machine translation systems, they can release themselves from various restrictions: restrictions of working site and the burden of hard work which demands that translators use their brains merely as part of their bodies." Mr.Furusho (Yumeya), who manages a personal computer network host station for sixty members, introduced a case of textbook preparation for English study by the auditory translation method.

### *Session 3*

The theme of the third session was "The potential of translation via network." In this session, Mr. Inoue (Iris International, Inc.) introduced the present state of machine translation service between English and several European languages via NIFTY-SERVE. Mr.Harada (Kodensha Co.Ltd.) introduced examples of Japanese-Korean translation system operation via personal computer networks. Mr. Kashiwabara (Subaru International Inc.) explained network services between machine translation systems and commercial databases, and between machine translation systems and contracted translators. Prof. Ishizaki (Keio Univ.) expressed his opinion as to the difficulty of system construction, giving an instance in which his students could not clearly distinguish machine translation from automatic translation.

### *Session 4*

In the fourth session, Mr. Yoshida, a lawyer, spoke on "Machine translation and copyright and intellectual property."

## **December 17: Symposium 2: Future Electronic Communication**

### *Session 1*

The theme of the first session was "Future electronic communication." Prof.Nishigaki (Meiji Univ.) gave the keynote speech, saying that toward the 21st century, international communication will be supported by electronic media, and that among those media, machine translation may play a vital role.

### *Session 2*

In the second session, a panel discussion by senior researchers was held under the theme "Future machine translation technique." Discussion was extremely vivid, presenting various opinions as to challenges in analysis technique and utilization of the pattern recognition technique, both of which machine translation must fully overcome; how to implant to machine systems the know-how that human translators have; how to incorporate tools to bridge the gap between translators and machine translation; opinions on the need to establish automatic pre-editing techniques to enhance translation efficiency; how to input common sense to computers, the range of common sense and so on.

### *Session 3*

In the third session, a panel discussion was held on "Future translation." Scholars of both science and literature participated, discussing from the cultural viewpoint how machine translation will affect translation work in the future. Some expressed their concerns that technical needs may deteriorate language, and eventually traditional culture, whether machine translation systems can convey human feelings, and whether or not machines force people to write dull sentences without the expression of

human emotions. The discussion was developed by the unique and lofty ideas that even though all new techniques evoke varied opposition and serious concern, most of which proves eventually to be beside the point, technological progress has always changed people and their environment, eventually contributing to the creation of a new culture.

### **Machine Translation Systems Showcase: December 15 and 16**

On December 15 and 16, an exhibition was held on machine translation systems, related software, and related tools. At this exhibition, three organizations and eleven enterprises displayed their merchandise.

In the booth of the CICC, researchers demonstrated evaluation tests of their systems to reveal the outcome of the past six-year international joint researches. The researchers who participated in this demonstration came from various countries: Thailand, Malaysia, Indonesia and China, as well as Japan. This machine translation system, employing the interlingual approach, can simultaneously translate several languages. Its current performance allows translation in 20 ways among 5 countries. Using international network lines, the system can convey, for instance, the speeches of delegates at the ASEAN meetings, in our own languages, automatically interpreted by the machine. Though this epoch-making system is still in the R & D stage, its early commercialization is expected. Translation demand is rapidly growing, commensurate with the globalization trend seen in the transfer of manufacturing facilities overseas, the establishment of overseas research and development centres, and the growing international information exchange.

To attract visitors to this exhibition, the organizers sent direct mail to the research industry, export/import industry, manufacturers of electric and information appliances, office apparatus, automobiles, and machines, which have relatively high overseas manufacturing ratios. During the two days of the exhibition, the venue was crowded with 3,500 visitors. Since the majority of the visitors had sizable translation demand in their business, they showed special interest in the displayed products, staying at the venue for a long time and asking questions eagerly. There were numerous people around the booths of several companies which had revealed new products immediately before the exhibition. The staff were exhausted giving explanations to hundreds of visitors, who seemed to have been impressed by the rapid progress of machine translation systems, and expressed their appreciation for the improving operability of products.

In conclusion, the Fair, attracting a large number of visitors, was successful even though there were several problems, including the short period for public information, the date of the fair falling in the busy year-end season, and poor access to the venue.

---

## **ASSOCIATION NEWS**

### **INTERNATIONAL ASSOCIATION FOR MACHINE TRANSLATION**

#### **Second General Assembly**

**Hotel Okura Kobe, Kobe, Japan**

**Thursday, 22 July 1993**

The Second General Assembly of IAMT was called to order by President Makoto Nagao at 4:40 p.m. and the provisional agenda was adopted.

President Nagao reported that since the creation of IAMT two years earlier at MT Summit III the membership of the three regions had grown to over 700 (615 individual and 86 corporate and institutional members). During this period the Association launched its newsletter, MT News International, under the editorship of John Hutchins, five issues of which had been published to date.

IAMT contributed financially to TMI-92 in Montreal and TMI-93 in Kyoto, and, in addition, it organized a workshop in collaboration with AMTA on the evaluation of MT, held in San Diego on November 1992.

The Secretary (Muriel Vasconcellos) reported that, pursuant to a decision of the 2nd meeting of the Council (Montreal, 25 June 1992), IAMT had been incorporated in Washington, D.C., as a non-profit association on 15 October 1992. The Bylaws, copies of which were circulated to the General Assembly, had been ratified by the Council at its 4th meeting on 21 July 1993. She explained that they provided for two new vice presidential positions in order to guarantee representation of all three regions on the Executive Committee. The two new positions are: Vice President for Programs, which is to be held by a representative of AAMT, and Vice President for Information, to be held by a representative of EAMT. Since the secretariat is in Washington, D.C., the positions of Secretary and Treasurer are filled by representatives of AMTA. The position of President-elect will rotate between the regions in tandem with that of President. Each region shall select its own representatives, including those to fill the respective offices, and the officers are confirmed by the General Assembly.

The financial report, also circulated to the Assembly, had been approved by the Council. As of 10 July there was a balance on hand of \$17,627.37. A budget for 1994 is in preparation.

Speaking for AAMT, Makoto Nagao reported that their journal, started three years ago by the Japan Association for Machine Translation, was now being published in both English and Japanese and circulated throughout the Asia-Pacific region. The Association held an "MT Fair" in December 1992 and organized a workshop on MT evaluation in June 1993. Its committees on evaluation and controlled language each conducted a study and prepared a monograph in their area of concern. AAMT had 200 individual members and 54 corporate or institutional members as of July 1993.

Reporting for the American region, Muriel Vasconcellos indicated that AMTA has built up a mailing list of about 2,500 names and has sent out several mailings. This list was enhanced by a large number of inquiries received following publication of a suite of articles on MT by AMTA authors in Byte magazine in January 1993, which included a box on AMTA and IAMT. As of the June there were 330 individual members and 17 corporate or institutional members.

The Regional Editor, Joseph Pentheroudakis, now does the page makeup for MT News International and arranges to print the copies for EAMT as well as AMTA. The Association has just published the MT Yellow Book (advance copies were made available at Summit IV for inspection), which includes a membership list and also serves as a locator of individuals, institutions, and companies involved in MT – mainly in the Americas. She hoped that it would grow to include the other regions as well.

AMTA organized a two-day workshop on MT evaluation (San Diego, 2-3 November 1992) and co-sponsored a seminar on translators and MT in cooperation with the American Translators Association (San Diego, 4 November). It also contributed US\$2,000 to TMI-92 in Montreal (25-27 June 1992). In 1993 AMTA will give a workshop on "Text and MT," dealing mainly with the translatability of MT input, and it will co-sponsor a seminar with ATA on "Matching Up MT Users and MT Environments," to be held in Philadelphia on 5 and 6 October, respectively. AMTA is considering having its own conference in 1994.

The activities of EAMT in the last two years, reported by Margaret King, included organization of a workshop on "Translation and Translation Theory," held immediately after Coling '92 (Nantes, 23-28 July) and workshops on "Computer-integrated Translation and the User" (Saarbrücken, 29-30 October 1992) and "The Lexicon and MT" (Heidelberg, 26-28 April 1993). In addition, EAMT will be lending its patronage to the Bulgarian Summer School on Computational Linguistics (5-11 September 1993), which this year emphasizes machine translation. At each event the opportunity is taken to hold a "rolling meeting" of EAMT members to consider MT-related issues. The EAMT held its first general assembly in Utrecht on 22 April 1993. It is now planning for its own three-day conference in 1994, one day each to be devoted to users, commercial systems, and research. There are 80 individual members and 15 corporate or institutional members.

At this point the following new officers of IAMT were presented for confirmation by the General Assembly: President, Margaret King; President-elect, Muriel Vasconcellos; and Secretary, W. Scott Bennett. Roberta Merchant continues as Treasurer. Outgoing President Nagao then passed the gavel to President-elect Margaret King, who took the chair. She stated that the aims of IAMT

will would not change and that the Association will continue to support MT users, developers, and researchers, as well as cooperation between them. She expressed admiration for the insightful leadership provided by Makoto Nagao as founder and first president of IAMT and thanked him for his hard work and enthusiasm. Finally, she appealed to members to come forward with new ideas.

John Hutchins presented a proposed resolution urging the regional associations to provide the editor of MTNI with copies of all their publications and requesting the organizers of meetings convened by the regional associations to prepare a report on such events for MTNI and to deposit a copy of the proceedings, if any, to all of the foregoing immediately upon publication. The proposal was seconded and approved unanimously as Resolution GA2/1 (see text below).

A second resolution was then presented by Margaret King, who explained that a member of EAMT had been made aware that a patent exists which might be used to discourage diversity and free competition in research, development, and the sale of machine translation systems. There was concern that the claims of the patent in this case were so general that they could, if they prevailed, affect all research and development in machine translation and natural language processing. In light of this threat to the science of machine translation, the proposed resolution was drafted opposing the inappropriate use of patents and other legal instruments for this purpose. She emphasized that the general principle was of greater concern than the particular case. There being a second and no objections, the proposal was approved as Resolution GA2/2 (see text below).

It was announced that the next General Assembly will take place at the time of MT Summit V, which will be held in Luxembourg on 11-13 July 1995. It was also announced that TMI-95 will be held in Louvain, Belgium.

The General Assembly was adjourned at 5:15 p.m.

Respectfully submitted,

Muriel Vasconcellos, Secretary

---

## **Resolutions of the Second General Assembly**

### Resolution GA2/1

#### "Material for MT News International"

The Second General Assembly of the International Association for Machine Translation,

Noting the importance of ensuring that MT News International offers complete coverage of the latest information in the field of machine translation; and

Concerned that unless steps are taken to guarantee the flow of information, the value of MT News International will not be improved or even maintained,

*resolves:*

To urge the IAMT regional associations:

1. To provide the Editor with copies of all publications produced by them or on their behalf immediately upon publication.

2. To call upon members involved in the organization of conferences, seminars, workshops held by, on behalf of, or supported by the regional associations to provide the Editor with a report of any such meetings within three weeks of the end of the meeting and with copies of any proceedings produced before, during, or after such meetings immediately upon publication.

3. To encourage all members to send copies or photocopies of all their publications to the Editor within one year of publication.

---

### Resolution GA2/2

#### "Diversity and Free Competition in Machine Translation"

The Second General Assembly of the International Association for Machine Translation,

Recognizing that diversity and health competition between machine translation systems is essential to their growth and improvement and to viability of the field of machine translation in general,

*declares:*

That it disapproves of the use of coercive measures of any kind, including the inappropriate use of patent and other legal instruments, to discourage diversity and free competition.

---

## **BYLAWS OF THE INTERNATIONAL ASSOCIATION FOR MACHINE TRANSLATION, INC. (IAMT)**

### ARTICLE 1. NAME AND GENERAL STRUCTURE

The name of this nonprofit association, a corporation organized and existing under the laws of the District of Columbia, United States of America, shall be the International Association for Machine Translation, Inc. (IAMT). Provision is made herein for the establishment of three regional divisions within IAMT.

### ARTICLE 2. PURPOSES

The purposes of the International Association for Machine Translation, Inc., shall be exclusively nonprofit, scientific, and educational. It shall bring together users, developers, researchers, sponsors, and other individuals or institutional or corporate entities interested in machine translation for the purpose of studying, evaluating, and understanding the science of machine translation and educating the public on important scientific techniques and principles of machine translation. In furtherance of these purposes, IAMT may carry out the following activities:

- a. Sharing knowledge about the science and technology of machine translation through the collection, compilation, exchange, and dissemination of information.
- b. Sponsoring and supporting workshops, symposia, and conferences on machine translation and related technologies and applications;
- c. Developing appropriate educational materials and programs;
- d. Facilitating access by researchers to machine-readable corpora and cooperating in the exchange of formats and text-encoding conventions; and
- e. Discussing and establishing reference criteria for the evaluation of machine translation technology.

### ARTICLE 3. MEMBERSHIP

Membership in IAMT shall be open to any person (individual or legal entity) interested in the purposes of IAMT, and members shall normally belong to one of the three regional associations corresponding to the regional divisions. Membership in a regional association automatically confers membership in IAMT. All matters relating to membership shall be decided by the regional divisions through their duly constituted associations, which shall contribute ten percent (10%) of their total membership dues to support the regular operations of IAMT. All members shall be entitled to participate in educational functions of IAMT and to receive educational materials distributed by IAMT.

### ARTICLE 4. REGIONAL DIVISIONS

IAMT shall have three regional divisions, as follows: the Asia-Pacific Region, the Americas Region, and the European Region. Each region shall have its own association, and it is hereby established that the regional associations shall be an integral part of IAMT.

### ARTICLE 5. ORGANS

The organs of IAMT shall be the General Assembly and the Council, which carries out the business of IAMT. The Council shall have an Executive Committee.

### ARTICLE 6. GENERAL ASSEMBLY

*Section 1. Sessions.* The General Assembly shall hold a regular session once every two years at the time of the biennial "MT Summit" conference, the site of which shall rotate regularly between the three regions of IAMT. Special sessions of the General Assembly may be convened by decision of the Council or by a written petition signed by ten (10) members of one of the three regional associations and submitted through that association, or five (5) members each of two of the three regional associations and so submitted. Notice of any regular or special session of the Assembly, together with a provisional agenda, shall be published in MT News International, the official publication of IAMT, at least sixty (60) days in advance, the date and place having been fixed by the Council.

*Section 2. Powers and Duties.* The General Assembly may adopt resolutions and approve recommendations to the Council, including proposed amendments to the present Bylaws. At each session it shall also hear and approve the minutes of the previous session; hear the reports of the President, the Secretary, and the Treasurer, including a current financial statement, and of the Editor and any committees that may have been appointed; and install any new officers or other members of the Council.

*Section 3. Quorum.* The General Assembly is entitled to deliberate validly regardless of the number of members present.

*Section 4. Voting and Proxy.* Decisions of the General Assembly shall require a two-thirds majority vote of members either present or represented by proxies bearing written powers.

#### ARTICLE 7. COUNCIL

*Section 1. Composition.* The Council shall consist of no less than ten (10) members, as follows: the President, the President-elect, the Vice President for Programs, the Vice President for Information, the Secretary, the Treasurer, the Editor-in-Chief of MT News International (hereinafter called "Editor"), and at least three other members. In addition, ad hoc non-voting members may be invited for the discussion of special activities or projects.

*Section 2. Regional Representation.* Each regional division shall have three positions on the Council at all times. The corresponding regional association shall designate, in accordance with its established legal arrangements, the individuals to serve in the positions allotted to it. Equitable representation of the regions shall be ensured in the following manner:

a. The office of President shall rotate every two years among the three regional divisions and shall correspond to the region in which the next MT Summit is to be held.

b. The office of President-elect shall rotate every two years among the three regional divisions and shall correspond to the region in which the MT Summit is to be held four years hence counting from the date on which the person takes office.

c. The office of Vice President for Programs shall be filled by the Asia-Pacific Division.

d. The office of Vice President for Information shall be filled by the European Division.

e. The offices of Secretary and Treasurer shall be filled by the Americas Division.

f. The Asia-Pacific Division and the European Division shall each have one more seat on the Council, and the region that is not currently represented in the position of either President or President-elect shall have an additional seat.

*Section 3. Editor.* The Editor of MT News International is appointed by the Council. He or she shall serve as a member of the Council and shall be entitled to vote on all matters except those relating to his or her appointment. The Council shall have the authority to pay an honorarium to the Editor.

*Section 4. Vacancies and Delegation of Representation.* In the event of a vacancy occurring on the Council, the corresponding regional association shall be responsible for filling it, in accordance with its established legal arrangements. If a member of the Council is unable to attend a regular or special meeting of the Council, the corresponding regional association shall delegate another representative to fill the seat in question on an interim basis.

*Section 5. Powers and Duties.* The Council shall have the power and authority to manage the property of IAMT and to regulate and govern its business and affairs; to determine policies and make changes therein; to amend the present Bylaws; to specify and review the work of the elected officers and the Editor; and to devise and carry out such measures as the membership may direct or which, in the judgment of the Council, are necessary or desirable on behalf of IAMT or in furtherance of its nonprofit purposes, objectives, and goals. It shall also have the power to dissolve IAMT.

*Section 6. Meetings.* The Council shall hold a meeting immediately prior to each General Assembly and, in addition, it shall meet whenever it is convened by the President because the business of IAMT so requires. The Secretary shall notify all members of the Council of the date and place of any regular or special meeting and provide them with the provisional agenda, such notice being sent by facsimile or other electronic means at least fifteen (15) days in advance of the meeting.

*Section 7. Quorum.* A quorum for holding a meeting of the Council shall consist of two members from each region, including the President, President-elect, or one of the Vice Presidents.

*Section 8. Voting.* The affirmative vote of the majority of such quorum, or of the number of members present and voting if more than a quorum is present, shall be the act of the Council and of IAMT, except that amendments to the present Bylaws and a decision to dissolve IAMT must be approved by all members of the Council.

*Section 9. Voting by Mail or Facsimile.* Any action may be taken by the Council by mail or facsimile ballot, provided that such action by the Council is approved in writing by two-thirds of the members of the Council. The written record of such action shall be kept on file by the Secretary.

#### ARTICLE 8. EXECUTIVE COMMITTEE

*Section 1. Composition.* The Executive Committee shall consist of the President, the President-elect, the Vice President for Programs, the Vice President for Information, the Secretary, and the Treasurer.

*Section 2. Powers and Duties.* The Executive Committee shall be charged with acting for and in the name of the Council in the intervals between its regular meetings. In such circumstances the Executive Committee shall have all the powers, duties, and attributes of the Council. All actions properly taken by this Committee shall be the actions of the Council and of IAMT and are not subject to ratification by the Council. Specifically, the Executive Committee may:

a. Advise the Council on matters of policy;

b. Approve financial transactions relating to IAMT activities for sums not to exceed an amount to be established by the Council;

c. Approve contracts and/or formal agreements between IAMT and other national and international public or private institutions which may be beneficial to the tax-exempt purposes of IAMT, involving sums not to exceed an amount to be established by the Council.

*Section 3. Meetings.* A meeting of the Executive Committee may be convened whenever the business of IAMT so requires. It may be called by the President, or in his or her absence, the President-elect, one or the other of whom shall preside at such meetings. Any action within the powers of this Committee may be discussed and voted on. Minutes of its

meetings shall be promptly circulated by the Secretary to all members of the Council. If there is objection to any decision of the Executive Committee, any action on that decision shall be delayed until a consensus can be reached by a two-thirds majority of the Council.

*Section 4. Quorum.* A quorum shall consist of three (3) members of the Executive Committee, one of whom must be either the President or the President-elect. The majority vote of the quorum, or of the members present if more than a quorum is present, shall be the act of the Executive Committee.

#### ARTICLE 9. DUTIES AND RESPONSIBILITIES OF THE OFFICERS

*Section 1. General.* The officers shall carry out those duties prescribed by law, by the Articles of Incorporation, these Bylaws, the General Assembly, and the Council.

*Section 2. President.* The President shall conduct all business and affairs as prescribed by law, by the Articles of Incorporation, these Bylaws, and the Council, and he or she shall represent IAMT in all matters. The President may delegate such representation to any of the Vice Presidents or such other eligible member of the Council as the President may select.

*Section 3. President-elect.* The President-elect shall assist the President in conducting the business of IAMT and shall perform such duties as may be assigned to him or her by the President or the Council.

*Section 4. Vice President for Programs.* The Vice President for Programs shall assist the President in planning, programming, and executing the various educational activities and projects of IAMT and shall be responsible for and represent the President in the coordination of these activities with the regional associations. This officer shall also perform such other duties as may be assigned to him or her by the President or the Council.

*Section 5. Vice President for Information.* The Vice President for Information shall assist the President in planning, programming, and executing activities in the information field and shall coordinate with the Editor on the preparation of MT News International. This Vice President shall also perform such other duties as may be assigned to him or her by the President or the Council.

*Section 6. Secretary.* The Secretary shall keep an accurate, continuous record, including historical documents, of meetings and any other proceedings of the officers, the Council, and the General Assembly, as well as reports of any committees that may be established. The Secretary shall also maintain a current accurate inventory of the assets of the Association other than the financial assets. The Secretary shall notify the membership at large, the members of the Council, and the Executive Committee, as appropriate, of regular and special sessions and meetings, as well as workshops and conferences, and he or she shall also perform such other duties as may be assigned to him or her by the President or the Council.

*Section 7. Treasurer.* The Treasurer shall:

a. Record and hold all funds of IAMT, including the ten percent (10%) contribution from membership dues collected by the regional associations, in such accounts as the Council may designate;

b. Submit to the Council a proposed budget for the next fiscal year at least sixty (60) days in advance and obtain the Council's approval thereof;

c. Disburse from the funds of IAMT expenditures duly authorized under the budget up to the amounts authorized;

d. When unexpected expenditures arise in excess of \$2,000, or when the cost of scheduled activities exceeds the budget by this amount, consult with the Executive Committee and obtain its approval prior to disbursement, and consult with the full Council in the event of expenditures in excess of \$5,000;

e. Keep accurate and complete records showing all receipts, deposits, and disbursements, allowing the books to be open at all reasonable times to members of the Council and the officers, inspectors, and auditors, and to any member having made a written request in advance;

f. Prepare the annual financial report and present it to the Council as soon as fiscal year ends for its approval, and also publish said financial report in MT News International and present it to the General Assembly; and

g. Perform such other duties as may be assigned by the President or the Council.

*Section 8. Compensation.* No compensation shall be paid to officers, members of the Council, or committee members for their duties related to these positions, except that the Editor of MT News International may be paid an honorarium.

#### ARTICLE 10. SECRETARIAT

The Secretariat of IAMT shall be located in Washington, D.C., in the United States of America.

#### ARTICLE 11. MT NEWS INTERNATIONAL

*Section 1. Name.* IAMT shall publish a regular newsletter known as MT News International.

*Section 2. Circulation.* MT News International shall be distributed to all members. It may also be sold on a subscription basis.

*Section 2. Editorial Board.* There shall be an Editorial Board composed of the Editor-in-Chief (referred to above as the "Editor"), one editor from each of the three regional associations, the Vice President for Information, and the President ex officio. The Editorial Board shall consult by mail or electronic means and may hold a meeting at the request of any two of its members.

*Section 3. Editorial Policy.* Editorial policy shall be set by the Council based on the recommendations of the Editorial Board.

*Section 4. Cost of Publication.* Editorial and setup costs associated with the preparation of MT News International shall be borne by IAMT; printing and distribution costs shall be borne by the respective regional associations as required.

#### ARTICLE 12. COMMITTEES

Standing and ad hoc committees may be appointed by the President or the Council. Their composition, terms of reference, powers, and duties shall be determined by the Council.

#### ARTICLE 13. PARLIAMENTARY PROCEDURE

The rules contained in Robert's Rules of Order Newly Revised shall govern IAMT proceedings in all cases to which they are applicable and in which they are not inconsistent with the Articles of Incorporation, these Bylaws, and any standing or special rules of order adopted by IAMT.

#### ARTICLE 14. AMENDMENTS

*Section 1. Submission and Approval of Proposed Amendments.* Amendments to the Articles of Incorporation or to these Bylaws may be proposed by a member of the Council or by a written petition signed by ten (10) members of one of the three regional associations and submitted through that association, or five (5) members each of two of the three regional associations and so submitted. Such amendments shall require the unanimous approval of the Council.

*Section 2. Effect.* Amendments that have been duly approved shall become effective as of the date they are approved. They shall be published in the next issue of MT News International and reported to the next General Assembly.

---

## USERS of SYSTEMS

### **The Current State of MT Usage – Or : How Do I Use Thee? Let Me Count the Ways<sup>1</sup>**

*Muriel Vasconcellos*

#### **PCMT: A New Passion that Changes Everything**

Two years ago, when we met in Washington at MT Summit III, it was obvious that MT was increasingly headed for the personal computer. Today the revolution is upon us. The advent of affordable software that can run on anyone's desktop ("PCMT"<sup>2</sup>) has totally challenged the received wisdom about MT usage. We must take a new look at the user profile, the purposes of MT, the products and the markets to which they are being directed, and the long-range future of the industry as a whole.

This report addresses the gap in our understanding of current MT usage by attempting an overview of all uses of MT based on the most concrete facts that could be found. It has considered only tried-and-true experiences and cumulative data reported directly by users. Information is particularly nebulous in the area of PCMT. Since there is no major up-front investment that needs to be justified, the user is less motivated to keep statistics. Nevertheless, some impressive facts are already a matter of record.

In the first place, there is now evidence that we are talking in rather large numbers of MT users. The June 1993 issue of *WordPerfect Magazine* reported the results of a mail-in poll in which readers voted for their favourite PCMT software. A total of 7,865 respondents took the trouble to send in their vote<sup>3</sup>. Presumably these people have road-tested at least one of the products and may in fact be using MT for practical purposes. The top three choices were Linguistic Products' PC-Translator, MicroTac Software's Translation Assistant, and Globalink's GTS (version unspecified). PC-Translator has doubled in sales each year since it first appeared on the market in 1985. The company periodically introduces improvements in its 12 language combinations and usually has new combinations in the pipeline; the developers have been heartened by the high percentage of registered users who request upgrades and new languages<sup>4</sup>. Globalink, which offers seven language combinations, went public in June 1993<sup>5</sup>, and their prospectus states that approximately 13,000 units have been sold or placed with dealers since 1990. MicroTac, for its part, leads the market by a wide margin: in May 1993, all-time total sales of its four bidirectional packages reached a staggering 150,000 units<sup>6</sup>. The Translation Assistants are priced at under US\$100 and, in some discount houses, as little as US\$60.



---

There are now over a dozen companies selling PCMT in the United States. Together they translate in a total of 17 different directions, and a number of other systems and language combinations are under development<sup>7</sup>.

These products are being used in myriad ways. In the long run, translation varies as greatly as the texts that undergo it, the people who perform the process, and the consumers who require it. Each use is somewhat unique.

Even more impressive than the numbers is the fact that many users of the PCMT systems are happy campers. Their ranks include both translators and nontranslators, and it is among the latter that PCMT is cutting its widest swath. From unsolicited testimonials received by the vendors<sup>8</sup>, we learn that many people are enlisting these packages to prepare letters and memos in languages that are foreign to them. One user of this kind writes: "The PC-Translator is doing wonderfully, we are all satisfied." There seems to be a slight preference for enlisting them to produce translations of texts prepared by the user rather than to comprehend foreign texts, which are typically input by hand or by a pesky process of optical scanning.

Sometimes the users do not know the target language at all. Installed on a laptop, PCMT has served as a practical companion in social situations where language is a barrier, and it has helped travellers to get around in foreign countries. An American in Paris reports that he used French Assistant to explain to the caretaker of his building that the hot water was off. Another MicroTac user, an American priest filling in at the last minute on a cruise ship, relied on this same software to prepare his sermon in French. Most touching, perhaps, is the user of Italian Assistant who wrote: "Through your product I have been able to correspond with my relatives in Italy since my trip in 1990, when I was introduced to them for the first time. My dad passed away two years ago and my mom is too old to write."

### **Finding the Real MT Users**

Finding out who really uses machine translation is no simple task. A few years ago it was possible, with help from the vendors, to identify at least those customers who were using MT on a significant scale. Today, however, with PCMT selling in large volume and with vendors busy attending to a broader customer base, the picture is far less clear. For the purpose of this report, a strategy was devised for locating a representative sample of MT users, who were to be presented with the following list of questions<sup>9</sup>.

### **Survey Questions**

*System used? Since when?*

*Language combinations (from → into)?*

*Hardware platform? Since when?*

*Form of input (e.g., disk, downloaded files, OCR, manual keying)?*

*Purpose of translation?*

*Type of documents translated -- discourse genre (e.g., "technical manuals"), subject matter?*

*Output per year (number of words) percentage of total translation volume?*

*Dictionary size (number of entries) for each language combination?*

*Description of personnel who use it (e.g., contract translators, etc.)? How many?*

*Type and amount of pre-editing done?*

*Type and amount of postediting done?*

*System for incorporating feedback from end-consumers?*

*Advantages, disadvantages of MT?*

*News flash: Latest developments? Novel uses of MT? Plans for the future?*

---

As the first step, a list was drawn up of known users for whom fax addresses were available<sup>10</sup>. There were 33 of these (two of whom could not be reached). Next, a list was prepared of individuals who had checked the "User" box on their application form when they joined the Association for Machine Translation in the Americas. This exercise garnered 15 more names. It was

known that some of these people were prospective users still investigating the feasibility of MT, so a letter was prepared addressing each one as a "user or potential user of MT" and asking them to report on their plans for using it if they did not already have it installed. The third step was to contact the vendors directly to ask them for the names and fax numbers of "some of [their] principal clients," sharing with them the list of questions that would be asked. Because of multiple sites and contacts, a total of 32 inquiries were sent out to vendors of 23 systems or families of systems. Six additional known vendors could not be reached. Of the 32 who were contacted, 14 replied and provided information about their users. These replies yielded 22 additional users, all of whom were approached. In the end, fax letters went out to 70 users or potential users.

Thus a fairly wide net was cast. Even so, the coverage was far from complete. The information obtained without the assistance of the vendors was not collected in any systematic way. In the vendor cycle, not all of them could be contacted, many who were contacted did not respond, and those who did reply did not necessarily give a full list of their customers. Response from the PCMT vendors, who account for far and away the largest volume of purchased (if not operating) units, was particularly low: only three replied, and only one of these directed us to specific users. Given such large gaps in the coverage, the answers received can only be considered representative of the vendors and users who were reached and had the time and inclination to share their experience. They do not speak for MT as a whole.

Another piece of missing information, which would be difficult for any survey to ferret out, is the user sites that have fallen by the wayside -- and why. This information is important for a full understanding of MT usage. However, it is hard to come by. One usually learns it by chance. Recently, for example, in a translation service that had shown positive results with MT, there was a breakdown in the hardware on which the system depends, and management was unwilling to buy the same equipment again. Elsewhere, an MT operation was eliminated because of a company-wide "reorganization" -- perhaps an indirect victim of the foundering economy. At yet other site the operation was dependent on an individual, and when that person left there is no structure to keep it going. There may also be MT failures in the true sense that the text was not a good match for the system or not enough time and money were being saved to justify the investment. For a variety of reasons, most of this information, which would be very illuminating, is kept dark.

Despite its limitations, however, the material collected for the present report is significant in many ways. Its very abundance gives it a certain authority. A total of 40 responses were received: 33 from actual MT users, one from a user with a commitment to start in July 1993, and six from companies that were in the process of investigating MT -- two were running pilot tests, one had put out an invitation to bid, and three were undertaking feasibility studies. CompuServe was included in this last group, with plans to offer on-line service from English to French starting in the fall of 1993 and other combinations later. In addition, answers to the same questions, gathered within the last nine months, were available from five other users and were included in the study. The analysis that follows covers the 33 responses from actual users and the five additional ones for which information was available, for a total of 38 user sites -- or 54% of those that had been contacted. In all, they represent 15 different systems: Atlas, DP/Translator, Duet Qt, Général TAO, Hicats, Shalt, JICST, Logos, MicroCat, Metal, PC-Translator, Pivot, NHK, Spanam/Engspan, and Systran (including Systran Express, the on-line service that anyone with a PC, a modem, and a checkbook can tap into). There were 16 users in the Americas, 11 from Europe, and 11 from Japan<sup>11</sup>. This may be the largest body of data ever collected at a single time on the use of MT. While it does not permit hard statistics, some very interesting conclusions can be drawn about how MT stands up to the test of translating texts in the real world.

### **Measuring MT Usage**

We can learn a lot about how much MT is being used from the volume of translation being produced and the percentage that this represents of the total workload. The survey yielded some illuminating information in this regard.

Thirty of the 38 users gave information on the volume of translation they produce using MT, the percentage that this represents of their total workload, or both (see table). Many of them had statistics at their fingertips, and it is easy to see that high-volume users, new or pilot users who are keeping a close watch on the effect of MT implementation, and users closely involved with development of the system itself would have reason to keep careful records.

In the category of large-volume users, the figures show that there are some truly industrial-strength MT operations. The European Commission is near the top of the list with 30 million words a year of general translation, for which they use Systran in a total of 13 different language combinations and serve from 400 to 500 end-consumers. These numbers take on special importance because the translations are in a wide range of subject areas and discourse genres. They amount to 15% of the total translation workload of the CEC. Interestingly, only 30% of Systran's output is postedited by professional translators; the rest is delivered "raw" and is used for information purposes only.

Two other very large users are Bull in France, which expects to be using Systran at an annual rate of 45 million words by the end of 1993, and Lexi-tech, which uses Logos for about 25 million words a year. Both these companies are using MT for technical documentation. Météo generates about 17 million words of weather bulletins each year for Environment Canada. The U.S. Air Force/FASTC, in its venerable information-gathering operation, annually translates between 10 and 12.5 million words with Systran. Intergraph relies on their own DP/Translator for about 10 million words. Xerox produces about 9 million words with Systran. Nikkei Printing uses NEC's Pivot and Sharp's Duet Qt for about 4.5 million words. And so on.

Added together, the volume of MT produced by these users -- slightly over half the known users approached in the survey -- comes to about 180 million words a year. MT use in the world undoubtedly exceeds 380 million. These figures translate, respectively, to some 720,000 pages of known use and about 1.2 million pages of estimated use. It is impossible to guess what percentage this represents of total translation in the world, however, since experts recognize that there is really no way to quantify the latter.

It can be seen from the table that the bulk of the work is translations of technical manuals and other material related to localization. The volume produced by the 15 users that provided this information comes to approximately 108 million, or 60% of the total volume reported. Of the entire sample of 38 users, 23, or 61% of them, fall into this category.

Another important parameter to look at is the proportion of the total translation load being handled by MT. The figures on percentage of the overall workload run the gamut. For the 24 who answered this question, the proportions ranged from 5% to 100% and formed an almost perfect bell-shaped curve. The average was 46% and the mean was 50%. Lexi-tech, one of the biggest users, relies on MT for 100% of its workload, and Nikkei Printing, also with a very large volume, uses it for 95%. Environment Canada uses Météo for 85% of all weather bulletins. The U.S. Air Force, which has had an MT installation since 1970, reports 80%. Some respondents seemed unclear on whether they should include languages not offered by their MT system in calculating the percentage, so it should be kept in mind that the figures may not always be referring to the same thing.

The high-percentage users are often high-volume users as well. The 10 respondents in the table that reported at least 50% usage and also reported figures for volume together produce 118.5 million words, or 66% of the total. As might be expected, many of these high-percentage users do technical manuals and other types of localization work: of the 12 users at 50% or higher, seven do this kind of work, and, as noted already, they account for a large share of the total volume. This should be concrete proof of the long-held assumption that there is a comfortable fit between technical manuals/localization and the automation of translation. In other words, MT does seem to work well for these applications.

Another interesting fact that emerges from the table is that most of the respondents have started using MT in the last five years. Of those in the table, 22, or 73%, began to use MT in 1988 or later. For the entire responding population of 38 users, the figure is 82%. In other words, MT use has

recently taken quite a spurt. And of course, with the advent of PCMT, this trend can be expected to accelerate sharply.

### **Contribution Required of the Human User**

Closely related to how much of the job MT is doing is the amount of human effort involved in the form of pre- and postediting. (None of the respondents had interactive workstations.)

*Pre-editing* was cited as a major issue only by the Raytheon user translating software written in the Ada programming language and by two of the three respondents who work with Japanese-English. While one said "pre-editing is basically division of long sentences and we usually don't spend that much time on it," another said that it is contracted out, and the third J-E user reported that pre-editing takes about 40% of total translation time. The other 30 respondents, all working with a Western language as the source, regarded pre-editing to be negligible or at least easily justifiable; 24 said they did little or none -- although interpretations of the term appeared to vary. Five said that they run an automatic spell-checker on the input; five mentioned conversion software or adaptation of the format; one referred to the need to proofread OCR output; and two indicated that pre-editing mainly involved blocking material that does not require translation. One user spends time "cutting overly long sentences into shorter ones, fixing up punctuation, etc."

A user in France has tried "end-user sensitization to 'clear writing,' with no evidence of success," while another one gives informal guidance on how to write for MT. Two said that their documents are written originally in a controlled language, and one reported that the input is edited to conform to the company's controlled language at a rate of 3,750 words a day -- which also happens to be their rate of *postediting*. Estimates of percentage of total translation time were given at 5%-10%, 10% (two respondents), and 20%-30%. One user included terminology research and dictionary maintenance under this heading, for approximately 60% of total MT time.

*Postediting*, on the other hand, generally accounted for a large share of production time and cost, and it was also the subject of a lot of comments when it came to discussing the disadvantages of MT. A number of respondents said that postediting is done directly on a word processor, one of them preferring commercial off-the-shelf word processing to the product developed by the MT vendor. Many pointed out that the requirement for postediting varies depending on the quality of the output, and that some language combinations give better results than others (e.g., "German-English [is better than] English-German"). The J-E user that did not report very much pre-editing said: "We rewrite the sentences after MT rather than [pre-]editing. Usually it takes a lot of time and manual power." An E-J user, in turn, felt that the main disadvantage of MT was the difficulty of postediting to achieve "acceptable" expressions in Japanese. The system developed by NHK has a user interface that presents several choices of output for the user to pick from, and the user can specify how many choices the system offers. Général TAO, when it gets overly challenged, leaves segments in the source language untranslated, and these passages must then be done by hand.

Several used the word "extensive" in characterizing their postediting. One respondent indicated that 75% or more of the text is touched during the postediting phase, although this proportion might vary depending on the translator, the product, or the language. On the other hand, Météo requires intervention in less than 5% of the output for a translation of good quality.

A number of respondents said that they review the entire text or do a "100% full postedit." This percentage should not be confused with the percentage of text that is actually corrected. A few require very high quality (e.g., for subtitles of television broadcasts, insurance contracts, publications), while some of them settle for an in-between product -- from "clean[ing] up the language, adjust[ing] the format, and review[ing] for technical accuracy," to "editing for accuracy but not for style unless requested," and, finally, to "quick and dirty." The U.S. Air Force has special software developed by Systran, called Editsys, which automatically picks out problem areas and leaves the rest of the text, usually about 80%, to be delivered without review. Some users have two levels of editing -- "information only" (or "for understanding only") versus a full translation. One respondent indicated that they offer both raw and reviewed translation but that only reviewed translation is "marketed" and accounts for 95% of their usage.

In terms of share of the total process, the user who said that terminology and dictionary work accounted for 60% of total MT time went on to attribute 20% of this time to postediting. Another said postediting represented 25% of the time. A third one said the proportion was 30%.

In the discussion of the disadvantages of MT, postediting kept coming up as a sore point. The respondents complained of the high cost, the time it takes, and the lack of user-friendly functions for posteditors.

### **"To the Level of Everyday's Most Quiet Need"**

Underlying the whole question of production is the purpose for which the translation is required. It is important to assess whether or not MT contributes to achieving the user's long-term service objective. As we saw earlier, a large percentage of the respondents are engaged in producing localization materials, often including immense volumes of technical manuals and, in at least three cases, software as well. Their responses definitely show that MT helps to move the process along so that they can get their products to market sooner. Perhaps the contribution of MT is not so much in producing a structurally correct text as it is in keeping terminology consistent and in eliminating the need to reintroduce graphics and format codes in target-language documents. Fisher-Rosemount, a high-volume user and manufacturer of machinery for industrial fluids, said that "translation would be barely feasible for this volume at this speed without it. By retaining formatting attributes, tables, and illustrations, [MT] saves enormous work and money." This user's bottom line: "Cost savings of nearly 50%." The sentiment is echoed by the owner of a commercial translation service that relies heavily on MT, who says: MT is "indispensable for high-volume jobs."

MT is being used for other purposes as well, of course. The sharing of scientific and technical information, especially from on-line databases, is a growing area. The U.S. Air Force (FASTC) has now expanded its MT operation to 17 subject fields and five languages and is starting to translate titles and short abstracts from on-line sources. Since 1990 the Japan Information Center for Science and Technology has been translating the mammoth JICST database into English with its own MT system and reports a 40% reduction in cost. Also in Japan, the Bio Information Center provides up-to-date data in medical and biotechnical fields (medical reports, database abstracts) with the help of MT, while the Pan American Health Organization in Washington, D.C., uses MT for publication-quality texts in similar technical fields as well as others. And Henkel KgaA in Düsseldorf uses MT to translate chemical abstracts, reports, and data sheets.

The Canadian agency DTSB-Statistics recently started using MT to translate technical papers and repetitive texts such as consumer price indexes for dissemination purposes. And of course Météo's weather bulletins for Environment Canada are a well-known example of MT use; translation is now bidirectional, and turnaround time for a given bulletin is less than 6 minutes.

One of the most novel uses of MT was reported at MT Summit III -- namely, NHK's television captioning project. Their MT system is now bundled in a prototype subtitle production system that also includes integrated modules for videotape monitoring on-screen, manual superimpose-timing input, and preview of the completed program. It was unveiled in June 1993.

From the users' responses, it would appear that the issue is not whether MT can meet these needs, but rather how efficiently it can do so. In some cases it has proved to be highly functional, while in others the jury is still out.

### **"With Smiles and Tears"**

The users were forthcoming about both the advantages and disadvantages of MT. Several listed a number of advantages and no disadvantages. The advantages cited most often were consistency of terminology, faster turnaround (to speed up market penetration), and increased productivity. One user commented that the terminology factor directly contributed to increased productivity ("at least 1.8 times better than human-only translation"). It was noted that certain types of errors are avoided -- e.g., skipped passages, numbers incorrectly copied. Filters on publishing systems which eliminate the need to re-enter format codes were very popular. Also cited was MT's ability to quickly process high volumes of material in many languages simultaneously.

Other specific comments were: "When the requester requires FYI translation, we can speed up the edit and still make the translation intelligible." "Less need for top quality translator." "We expect a capacity increase as soon as we have gained more experience with the system" (a user who started at beginning of 1993). "It gets better" (a new user).

And from the operator's perspective: "Lightens the translator's load." "No cumbersome typing." "It also maintains the original format created in WordPerfect." "Beneficial for us because the kind of text we translate is very dry and very repetitive." "I really enjoy working with DP/Translator; it requires a lot of work at the beginning with the creation of custom dictionaries but helps maintain consistency. The machine generates a draft translation, performing the most boring part of the task, so that I can concentrate on perfecting the output."

The respondents were equally expressive about the disadvantages. Many of them complained about the poor quality of the output and the cumbersome process of postediting. They want better interfaces and postediting tools.

From the manager's viewpoint, several respondents cited the high cost of source text preparation and postediting. Two said it was difficult to find texts suitable for MT. One complained that it involves a lot of training, and two of them noted that it's costly for smaller projects. Another remarked that system development is too slow and that there should be more user support. In one case it was noted that inclusion of MT in the production scheme had complicated the workflow. With regard to one particular system, the respondent mentioned that enhancements are very costly because of its size. Two of them regretted that hardcopy input documents were not scannable; "efficiency from the use of MT is largely lost in the time required to manually key in a text." A user of the old Weidner MicroCat workstation reported that the equipment is wearing out and the alternatives seem too expensive. Also cited were the high cost of purchase and maintenance; complicated handling; "an un-ergonomic user interface"; lack of acceptance by internal translators. A new user said: "No improvement in speed so far."

Other comments were: "It somewhat inhibits creativity"; "loss of idiomaticity and style"; "resulting text is a little stilted and awkward"; "excessive adherence to MT output changes expression"; "translation system not sufficiently flexible about using one term in one context but another in a different context."

The following response gave real food for thought: "Up to now we have not really been able to make use of the advantages (consistency of terminology, speed, etc.). One of the advantages mentioned by salesmen, etc., [namely] that MT relieves translators of boring, repetitive tasks, is not relevant in my opinion as there are other repetitive tasks instead: text conversion, parameter editing, reformatting, writing Pattern Matcher instructions, reformatting, etc. I enjoy working with MT because it is an interesting tool and you learn a lot, but whether it really beats manual translation remains to be seen."

### **The "Future's Epigraph"**

By and large the users have a positive outlook, a desire to streamline their MT operations, and a keen interest in introducing improvements and trying out new applications. One current user plans to take on a new application, joining the ranks of those who use MT to screen translation requests. Another site is plugging MT into databases on CD-ROM.

They are asking for, and working on, new and better tools. They want to be on high-end workstations instead of mainframes. They want software to test texts ahead of time to see if they lend themselves to MT. Much in demand is a good system for repetitions processing, whereby previously translated texts are matched against the ongoing translation process and displayed for possible pasting in. They need better converters for moving freely between different publishing environments. They are also working on terminology managers. Integration of the workstation seems to be the key. The Canadian Government is putting the finishing touches on a "fully equipped zero-wait-time multimedia workstation on a LAN server" with access to terminology banks, multi-task word-processing packages, automated terminology searching, text analysis, and other specialized software.

They are also asking for, and working on, more language combinations, more domains, and better strategies for controlling the quality of input texts. At least two of them are seriously looking into interlingual MT, and the Unión Fenosa in Spain, working with Carnegie Mellon's Kant system, is dreaming the impossible dream and turning it into reality: MT with no postediting!

## Notes

1. Updated version of invited lecture presented at MT Summit IV (Kobe, Japan, 19-22 July 1993). Published with permission.
2. "PCMT" is understood here to refer to PC-based MT products that do full-sentence batch translation.
3. From a larger "Reader's Choice" questionnaire, this number of people cast votes specifically for a PCMT package (source: Shannon Harmon, WordPerfect Corporation).
4. Source: Ralph Dessau, Linguistic Products.
5. GLNK U on the National Capitalization Market. Globalink regretted not being able to provide more information for the current report but was under a routine temporary period of silence.
6. Source: Michael Tacelosky, President, MicroTac Software (figure does not include upgrades.)
7. Source: "Report on PC-based MT products", American Translators Association, December 1992, compiled by L.Chris Miller.
8. Copies of the original testimonials provided by Linguistic Products and MicroTac Software.
9. Questions based on a model developed by Joann Ryan for research presented at the seminar "Machine Translation for Translators" (San Diego, 4 November 1992), sponsored jointly by the American Translators Association and the Association for Machine Translation in the Americas.
10. The entry criterion for the study was that the user could be reached by fax.
11. The list of users, together with the type of text they are translating, was published in the proceedings of MT Summit IV. To this list should be added late responses received from the Commission of the European Communities (general and technical translation, including 70% information-only), Inter Group (technical manuals), and JAPO (patent titles and abstracts), which are included in the totals cited in the present version of the paper.

---

### Summary of MT Use by Survey Respondents<sup>a</sup>

User #	Year of startup	Estimated no. of words per year <sup>b</sup> (thousands)	Percentage of total volume	Type of text
1	1970	11,250	80	Scientific and technical articles
2	1977	17,000	85	Weather bulletins
3	1978	9,000	50	Dissemination
4	1980	2,500	67	General and technical
5	1981	30,000	15	Low-level in-house documents
6	1982	-	10	Technical manuals
7	1986	10-100	100	Service publications
8	1987	-	20	Technical manuals
9	1988	25,000	100	Technical manuals
10	1988	10,000	-	Software, hardware documentation
11	1988	4,500	95	Technical manuals
12	1988	1,600	-	Technical manuals
13	1988	-	10	Customer documentation
14	1989	2,500-3,000	40	Technical manuals
15	1989	44-60	-	Subtitles for news in English
16	1989	750-1,000	5	Internal technical documentation
17	1990	2,500	50	Insurance and pension contracts
18	1990	3,445 <sup>c</sup>	- <sup>e</sup>	Titles + abstracts, JICST database
19	1990	2,000	25	On-line, hardcopy documentation
20	1990	480	-	Technical manuals
21	1990	350	20	Technical manuals
22	1991	1,600	67	Technical manuals

23	1991	375	30	Manuals, technical reports
24	1991	-	80	Chemical abstracts, data sheets
25	1992	45,000	50	Technical manuals
26	1992	1,500	-	Software, user manuals
27	1992	345 <sup>d</sup>	9	Titles of unexamined patents
28	1992	25	5	Scientific publications
29	1993	3,300	30	Technical manuals, price indexes
30	1993	-	90	Computer manuals

a. Eight of the 38 respondents did not provide the information being compared in this table.

b. Figures for numbers of pages were multiplied by 250 to permit comparison. Those for less than a year were annualized.

c. 85,000 titles plus 15,000 abstracts; average length of title estimated at 10 English words and average length of abstract (200 Japanese characters with upper limit of 300) estimated at 150 English words.

d. About 23,000 titles per year at an average of 15 English words each.

e. 90% of the abstract are written in English by bilingual abstractors; of the remaining 10%, all (100%) are translated by MT.

## Survey of MT usage in Japan Report of the System Application Technic Work Group of AAMT

*Takenori Makino*

*[Edited extract from: AAMT Journal, no.3 May 1993, p.1-5]*

### Use of MT

The analysis of answers to the questionnaire revealed the following facts. The large-scale user who uses the MT system systematically translates documents of a few hundred to a thousand pages per one MT terminal a month. Their target is rather limited, e.g. manuals, technical documents, patents, etc. so translation efficiency gets better.

On the other hand, small-scale users, such as personal users, translate only a few pages a month. Many of them use MT for raw translation of manuals and rough reading of technical documents.

Time spent on pre-editing ranges from a few minutes to fifteen minutes per page (A4 letter size), while postediting takes longer - 10 to 30 minutes. Most pre-editing is done for Japanese input. For example, long sentences are paraphrased as short ones, and omitted elements in sentences are supplied to make translation easier. The main tasks of postediting are the replacement of translated words and the refinement of the output... It seems that many successful users have expertise in editing and dictionary management, although these circumstances depend on the nature of their system. Some users have developed original support tools for MT systems...

Generally, the people working with MT systems are classified as pre-editors, post-editors, rewriters and operators. For small-scale users, one person fills all roles. Even with large-scale users, roles are not clearly separated. It was also found that English abilities are not always equal to those of English specialists (1st class of English examination in Japan). It means that ability to use an MT system is considered more important than personal ability in English. Almost all users train MT workers in their own company or office.

Users of Japanese-English systems are particularly insistent on improvements in translation quality. However, large-scale users with experience of actual use want easier operability and compatibility with other MT systems and software. This is because their target is to raise the efficiency of the MT system or of the total translation operation from an understanding of the translation capability of the present system.

### User dictionaries



The number of responses to the questionnaire was 31. The major fields are information, communication, electronics, machinery and aviation. The sizes of the Japanese-English and English-Japanese dictionaries are nearly the same. Many dictionaries have about 2,000-6,000 words. In the case of manuals, user dictionaries are often made for specific types of machines or airplanes. Most entries are for nouns. Parts of speech and semantic markers are included, but not the meaning of a word or an example sentence. However, many of the dictionaries stored in spreadsheets, RDB or text files have information about synonyms, explanation of word meanings and example sentences. Many answers indicated that priority of translation equivalents is made according to frequency of use. But human judgement is necessary for final selection...

The following were mentioned as problems in use: entry of verbs and idioms; priority between user dictionaries and basic dictionaries; timing for revision; handling compound names (in chemistry); unifying items and formats of dictionaries between manufacturers. Few users comprehend correctly statistical data about numbers of words, costs of dictionary compilation, and so on...

### **Tentative conclusions**

There is no way to decrease translation costs dramatically other than by using MT. The greatest motivation for introducing MT is to decrease translation costs. The other motivation is pure interest in MT as such. This survey, however, shows that many users do not use MT systems effectively in practice. The effective users of MT systems are companies that have to translate large volumes of documents: document translation companies, translation divisions in large companies, and so on.

The characteristics of effective users are:

- 1) they manage translation processes and personnel efficiently
- 2) they have enough experience of MT to stand the learning curve
- 3) they have examined the capability of MT, and they use MT systems within their capabilities.
- 4) they have trained their translation engineer within their organization.

In addition, they have reviewed their MT use as follows:

- 1) they construct a working environment in which resources such as dictionaries are shared among users
- 2) they have integrated computerized document processing
- 3) they have examined various methods of pre-editing

We believe it is important to share such experience among MT users in order to increasingly popularize MT systems.

The result of this survey is planned to be published soon. Please contact the editor [Hirosato Nomura, address on front page of MTNI] about its availability in English. The report will be combined with that conducted by another working group "MT System Evaluation Committee".

---

## **SYSTEMS AND PROJECTS**

### **CSK introduces ARGO J/E**

*[From: AAMT Journal, no.2 February 1993, p.6]*

Developed and enhanced over six years of R&D activity along with actual operation for the "Japan News Retrieval" service [translation service for news and articles of money market and securities over the Nikkei Telecom network], ARGO J/E, a Japanese-English machine translation software package, promises to give a new key to Japan's MT market.

ARGO J/E, which was formally unveiled by CSK (F16 Shinjuku Sumitomo Bldg., Nishi-Shinjuku 2-6-1, Shinjuku-ku, Tokyo, Japan) on December 11, 1992, uses the "modification detection system" that is a CSK-specific translation method. [This is a] syntactic analysis method by which the

words in a sentence are analyzed to detect the "modification relationship" between them (or to build a syntactic-semantic structure or tree of them). Parts of speech, semantic primitives, cases, and attributes are used for the establishment of a "modify-modified" structure. It permits analysis of sentences written in the Japanese language which otherwise could not be achieved because of their relatively free word order. The current version of ARGO J/E, features that CSK staff having experience in both development and internal trial operation will, on behalf of each user, take charge of building a special dictionary containing terms and idioms of his own. It will prove a powerful translation support, claims CSK, by clearing most MT users of their protesting complaint, "I have it, but always fail to manage it."

ARGO J/E is open to users. Its basic concept is that, as the user gets accustomed to his system, it should be growing to meet his requirements. The multiwindow-based graphical user interface (GUI) and the user terminology building tool enable the user to add new nouns, verbs and idioms, and special semantics between words to his dictionary database.

ARGO J/E is available with optional dictionaries of technical terminology including geography, biography, and information processing. The user can use more than one of those dictionaries and user dictionaries simultaneously so that he can translate sentences describing information of different fields at a time.

The price of an ARGO J/E package, 5.6 million yen, consists of 2.0 for the software and 3.6 for the organization of a user dictionary by CSK staff.

CSK Marketing is planning to supply both Japanese and overseas markets with 2,000 sets of ARGO J/E in five years over Nihon Sun Microsystems' sales network.

ARGO J/E is designed to work in the following hardware and software environments:

*Hardware:* SUN SPARCstation2 (main memory, 32 MB; swap area, 100 MB; hard disk, 300 MB minimum)

*Software:* SUN OS 4.1.1 (OS); OpenWindows 2.0.1-up (window environment); JLE 1.1-up (Japanese language environment)

For further information call +81-3-3342-3047 CSK Machine Translation Development Dept. (c/o Mr. Nagano)

---

## LogoVista E to J

[From AAMT Journal, no.2 February 1993, p.9-10]

### Outline of LogoVista E to J

As the internationalization of Japan has progressed, rapid and accurate translation of large volumes of English-language information has become increasingly important for those at the forefront of the business world.

LogoVista E to J is an English-to-Japanese translation support system based on the theories of Professor Susumu Kuno, an internationally recognized linguistics scholar.

LogoVista E to J features "high accuracy in syntactic parsing of the input English sentences," "appropriate semantic processing," and "sophisticated synthesis of Japanese sentences." It has been well received as a system that produces "usable translations."

The system's dictionary component, one of the keys to translation accuracy, consists of a Main Dictionary, used in conjunction with high-precision grammar rules for parsing; Technical Dictionaries containing specialized terms used in business and a variety of scientific and technical fields; and a User Dictionary, which individual users can build to fit their needs.

LogoVista E to J has been designed as an "open system" that is not restricted to one platform. It can be used on a variety of workstations and personal computers.

### Outline of LogoVista E to J's Linguistic Characteristics

The quality of the translation output is the most important factor in the overall quality of a translation support system. However, a lack of standard criteria impedes the evaluation of the output of

translation systems. We will illustrate the linguistic characteristics of LogoVista E to J by using real examples of its translation output. [Omitted - see original article.]

### **Major Features of the System**

The workstation and personal computer versions of LogoVista E to J have been designed to have identical features. You can use the system on a platform appropriate for the volume of your translation needs.

a) *Ease of Use.* LogoVista E to J is operated by using a mouse to select commands from pull-down menus. The simplicity of the design allows people with little knowledge of computers to use the system.

b) *Overview of the Translation Process Text Input.* Users can directly type in text to be translated, or an OCR system can be used when large volumes of text need to be input.

### **Translation Options**

Users can control the following translation options:

- \* the maximum number of Alternate Translations (from 2 to 20)
- \* the formality of the Japanese output (*da*, *dearu*, or *desu*)
- \* output of proper names (Japanese only, English only, or Japanese with English in parentheses)
- \* translation of definite articles (*sono*)
- \* input text type (dialogue or standard)
- \* text layout (Side-by-side or Formatted)
- \* use of Technical Dictionaries (several dictionaries can be used at the same time)

### **The Translation Process**

The input text is analyzed and translated. LogoVista E to J's progress in translating each sentence is displayed on the screen, as is the system's progress through the entire document being translated. When the processing of each sentence is completed, its translation is displayed on the screen.

### **Postediting**

The user can produce a more appropriate translation by using the Alternate Translation function.

The Alternate Translation function allows the user to choose an alternative translation of an entire sentence or of a selected word or phrase.

LogoVista E to J "learns" statistically from the choices made by the user while postediting with the Alternate Translation feature. In this way, the system is customized as it is used.

The user can also edit the translation output directly without using the Alternate Translation function.

The user can further customize LogoVista E to J by building a User Dictionary (including nouns, verbs, adjectives, and adverbs) created through simple, automated procedures.

For inquiries, please contact: LogoVista Corporation, 10-24, Shiomi 2-chome, Koto-ku, Tokyo 135 Japan. Tel: +81-3-5690-8531; Fax: +81-3-5690-1290

---

**Oki Electric Industry Co., Ltd. PENSÉE-GV**

*[From: AAMT Journal, no.3 May 1993, p.16-17]*

PENSÉE-GV is a Japanese-English/English-Japanese translation software developed by Oki Electric Industry Co.Ltd. Main advantage of PENSÉE-GV is to translate documents which may include figures and tables, keeping the original layout. PENSÉE-GV is called from GlobalView, an integrated document processing software developed by Fuji Xerox Corp.

Previous machine translation systems dealt with only texts. So, the user must have extracted texts from the documents beforehand, and returned the texts into the original places after the translation. This sequence of work included time-consuming cut-and-paste and re-formatting. These were routine activities and hoped to be eliminated.

Oki's policy for the machine translation system is to remove as many routine activities as possible, which will lead to inefficiency.

Based on the policy, Oki developed a JE/EJ translation support software, PENSÉE-GV, which translates documents into target language keeping the original layout.

### ***Features***

#### **\* Easy to use**

The user just clicks target source document icon(s) and selects EJ or JE translation operation in the pull-down menu, the PENSÉE automatically translates the source document(s) into the target language and shows the translated document icon(s). Just selecting the target word(s), sentence(s) and folder(s), the operator also easily choose the various modes from word-by-word translation to folder (collection of documents) translation.

#### **\* Preservation of format**

The translation retains the layout and font information of the original pages even with figures and lists. Reformatting and re-inputting become unnecessary.

#### **\* High-quality translation**

We achieve high-quality translation based on our 'Deep Case Grammar', which is a refinement of Case Grammar by our AI technology. This enables PENSÉE to capture deep structure of long and complicated sentences.

#### **\* High-speed translation**

PENSÉE-GV translates about 15,000 words per hour. It takes only tens of seconds for the 1 page translation.

#### **\* Options of translation**

In JE translation, PENSÉE-GV has the alternative translation modes for non-subject sentences, which frequently appear in manuals: (1) to translate them in the passive voice (standard mode); (2) to translate them in the imperative (MEIREI mode).

In EJ translation, PENSÉE-GV has the alternative translation modes for the Japanese ending postpositional particle: (1) to use 'DEARU' way (standard mode); to use 'DESU' or 'MASU' way (KEITAI mode).

### ***Specifications***

Translation languages:

Japanese/English, English/Japanese

Desk Top Publishing software:

GlobalView

Translation engine: PENSÉE

Hardware: OKITAC-S

Memory required: 16 MB or more (recommended 32 MB or more)

Hard disk required: 400 MB or more (including OS and GlobalView)

OS etc.: SUN OS 4.1.1 or later

+ JLE 1.1.1 or later

+ OS/FX optional function

+ GlobalView

System dictionary:

JE approximately 90,000 words

EJ approximately 60,000 words

Special dictionary (optional):

JE approximately 50,000 words (16 fields)

EJ approximately 50,000 words (16 fields)

(information processing, automobile, metal, mathematics, etc.)

User dictionary: unlimited within the capacity of the disk

Translation speed: approx. 15,000 words/hour (differ by machine)

Translation option:

JE: standard/MEIREI

EJ: standard/KEITAI

OCR (optional):

Japanese: SPARC Reader

English: ScanWorX

Readable document forms: plain text, JStar, ICHITARO, OASYS, Lotus 123, etc.

(Conversion software is required except for plain text and JStar. Some figures can not be read.)

**Contact.** For more detailed information, please inquire at: Oki Electric Industry Co.Ltd. Tel: +81-3-3454-2111

---

## DP/Translator Users Group Meets in Huntsville

Users of Intergraph Corporation's DP/Translator MT system met for the first time in Huntsville, Alabama, on 4 and 5 May of this year. Four sessions dealing with MT were held during the annual meeting of the Intergraph Graphics Users Group, which is the worldwide organization for users of Intergraph products.

On the first day, the DP/Translator product development group gave presentations on the new features of version 1.4 just released, and on future product direction. The role of user feedback in development was emphasized, along with the means by which it can be provided through the customer hotline, a database tracking system for problem reports, and the Special Interest Group organization.

A panel of users convened on the second day to discuss their implementations of a number of DP/Translator language pairs in different translation environments. Users from corporate translation departments as well as a translation service bureau participated. The type of translation being done varied from information-only translation for internal use to the translation of user documentation for publication.

One area of much discussion concerned typical problems in MT output. Guidelines for writers and translators in avoiding these problems were a topic of particular interest. A report on a joint experiment by technical writers and translators indicated that adhering to writing guidelines that allowed for better machine translation output also improved the quality of the original document. The results were seen as significant for organizations that can exercise control over the writing process and important in achieving cooperation between writers and translators. Other users had also observed that the quality of the computer translation varied with documents written by different writers or in different styles.

Productivity was another important topic of discussion. One user provided some recent statistics on translator productivity when employing DP/Translator to translate scientific articles on request for information only. After 3 months of concentrated work in the lexicon, productivity gains were beginning to be realized, and client acceptance of lightly edited computer output was favorable. With no control over the input for translation, building large dictionaries was seen as the most important factor in improving translation quality and increasing productivity in the post-editing phase. Comparisons of translator productivity in terms of words per day of finished translation seemed difficult, however, since in each organization translators had different duties.

Also on the second day was a presentation on using DP/Translator in a translation service bureau. Mentioned as key requirements in using MT successfully were the development large lexicons, having the ability to accept documents from clients using many different publishing applications, and having translators experienced in MT post-editing. Business growth and increased

hiring of translators underscore the success of MT in the service bureau environment when properly managed.

For further information about DP/Translator please contact Gary Thornton, Electronic Publishing Division, Intergraph Corporation, Mail Stop LR23A2, Huntsville, Alabama 35894-0001, USA.

---

## News from the Confederation of Independent States

*Evgeny Lovtsky*

An international festival "Language Education Culture" was held in Moscow on 23rd to 25th April in the Manège... The festival was dedicated to computer methods in education and cultural exchange, but the VCP ['All-Union Centre for Translation'] managed to demonstrate some of their NLP products (in addition to appropriate language-teaching programmes).

One was B. Tikhomirov's PC-based English-into-Russian MT system SPRINT. Based on their earlier mainframe version AMPAR which supported only batch processing, the system has acquired new features exploiting the possibilities offered by personal computers. For example, when editing a translation, the user can request a list of synonyms for any word of the text and make a desirable substitution by pressing a couple of keys. Essentially, however, the system is the same old AMPAR with all its merits and shortcomings. It is very fast and has large dictionaries, but due to some intrinsic inconsistencies the general quality of translation will not rise above some critical threshold.

The second was a Russian-into-English MT system ASPERA developed under the direction of Edward Korolev, a 1967 graduate of the Moscow Institute of Foreign Languages, Department of Applied Linguistics. The system is operational, which is a notable achievement and can be considered a success considering the conditions in which we all work. It belongs to dictionary-type systems with most of the linguistic information kept in the dictionary. To save time on development, they use off-the-shelf FoxBase software for storing the dictionary in computer memory. It may be a good idea for a prototype, but the inevitable deterioration of such parameters as word-per-second translation rate and hard disk memory demand will eventually impair its competitiveness.

I do not like the notation they use - random strings of letters instead of generally accepted abbreviations. It may seem unimportant, but prospective users willing to update dictionaries may find it inconvenient. The semantic component comprises about one hundred semantic primitives; I fear this number may be difficult to handle not only for an outsider, but even for a system linguist. Anyhow, these shortcomings do not belittle the importance of their contribution to machine translation.

[Some time after the festival] I met Michael Blekhman ... [who] had brought along his two-way English-Russian MT system PARS... From the point of view of its software organization, the system is quite user-friendly; at any point one can get a pop-up prompt by pressing the "HELP" key. The arrangement of menus and work screens is convenient. Both sentence-by-sentence and batch translation modes are possible. There is a built-in customized word processor. However, linguistically, the system is rather simple, to say the least. All grammar rules are directly programmed in PASCAL, which makes them hard to correct and sets rigid limits on the system's sophistication. The morphological information in the dictionary is exhaustive, but the syntactical one is scarce, and semantic information is practically non-existent. No matter how many times you enter a Russian word in the dictionary, you must each time set off its root and specify its declination or conjugation. As I understood, Misha and his co-workers are developing a new, improved version of their systems which will be free of those drawbacks.

[The first report by Evgeny Lovtsky on MT in the CIS appeared in MTNI#5, p.8-9.]

---

## Exhibitions at the MT Summit: brief report

Seventeen exhibitors were displaying their products and systems in two rooms in the Hotel Okura during the MT Summit in Kobe.

A number of manufacturers demonstrated the latest versions of their software. Toshiba was showing the ASTRANSAC system for Japanese-English translation, which is now available on Sun Workstations. At the NEC stand could be seen a demonstration of the well-established PIVOT system in its Japanese-English version, now incorporating a learning mechanism and pattern matching components. Nippon Telegraph and Telephone Corporation was exhibiting ALT/JE. The latest versions of the HICATS Japanese-English and English-Japanese systems were on display at the Hitachi stand. Sharp Corporation was demonstrating its successful Duet-Qt, with Japanese-English and English-Japanese on the same workstation. The ATLAS bi-directional (English/Japanese) system was on display at the Fujitsu stand together with the same company's ReadyPen aid for Japanese writing English business letters, with a database of sentence examples and grammar checkers. Nova Inc. was showing its PC-Transer/EJ (shortly to be available on the Macintosh) and the newly released Japanese-English version. Finally there were two separate English-Japanese systems from Catena: the familiar UNIX-based STAR system (now also for Macintosh with the name The Translator), and one of the newest systems on the market - from the LogoVista Corporation, which is part owned by Catena: this is LogoVista E to J (marketed in the United States by Language Engineering Corporation.) [For more on this system see elsewhere in this issue.]

Other systems on display are not commercial products. Matsushita was showing a prototype of its Japanese-English machine aided system, incorporating the latest component, the Modification Deciding Network. NHK was demonstrating its in-house English-Japanese system for translating foreign news articles. From Korea there was a demonstration of KAIST's prototype MATES/EK system for English-Korean translation in the area of electronics and electrical engineering. The government-supported multilingual CICC project was showing the prototype of its interlingua-based system intended for translation between the Japanese, Chinese, Thai, Malay and Indonesian languages. At a nearby stand, there was a demonstration of the EDR project; the Electronic Dictionary for Japanese and English is now in the final stages, and its sophisticated interfaces and access facilities were on display. And next to it was the ATR demonstration of ASURA, the latest version of the spoken language telephone interpreter, incorporating speech recognition, MT and speech synthesis within the domain of conference registration and hotel booking. On display was the Japanese speech recognizer and MT into English text on screen.

The only non-Japanese exhibitor was the Canadian Workplace Automation Center (now CITI). They were showing the prototype of the TransSearch software which enables translators access to a 'translation memory' of previously translated texts, with rapid display of text segments containing translations of words or phrases encountered in source texts.

Finally, Ricoh demonstrated its IMAGIO MF530 Dictionary Copier, which adds Japanese dictionary translations of a selection of English words to a photocopy of the original English text. Output can be either interlinear or as lists of English and Japanese words. The machine could well have attractions for the layman as the often-mooted 'translator-copier'.

---

## **RESEARCH DEVELOPMENTS**

### **PANGLOSS at the Computing Research Laboratory (NMSU) with Latest Evaluation Results**

*David Farwell, Steve Helmreich, Louise Guthrie and Yorick Wilks*

To many people's surprise, the US Department of Defense two years ago began a major program to support machine translation again after a gap of nearly thirty years. Although the original request for

proposals specified machine-aided translation, they finally decided to support three very different projects: a hybrid symbolic-statistical system at Dragon Systems; a basically statistically-driven system at IBM Research, and PANGLOSS, the basically symbolic system described below [and in the last MTNI issue's interview with Ed Hovy at ISI]. Since the systems are so different in methodology and in the amount of human interaction they assume, it has proved difficult to devise an evaluation regime to compare them, although that is what all ARPA programs normally do with competing groups they fund. Much energy and ingenuity has gone into such an evaluation regime though as yet any comparisons between the three systems are hard to make. The results of the most recent evaluation cycle are just in.

The Computing Research Laboratory (CRL) at New Mexico State University, jointly with the Center for Machine Translation (CMT) at Carnegie Mellon University and the Information Sciences Institute (ISI) at the University of Southern California, set out to develop a Translator's Workstation to assist a user in the translation of newspaper articles in the area of finance (mergers and acquisitions, followed by joint ventures) in one language (Spanish initially, followed by Japanese) into a second language (English). This project was dubbed PANGLOSS not simply due to its interlingual nature (PAN-GLOSS = all tongues), but to distinguish it from the CANDIDE system at IBM. Dr. Pangloss, of course, was Candide's teacher in Voltaire's picaresque novel. At the core of the project is a multi-language, knowledge-based, interlingual, interactive, machine-assisted translation system consisting of a source language analysis component, an interactive augmentor, and a target language generation component. The augmentor carries out a procedure which is partly automatic and not to be confused with postediting. In principle augmentation can be thought of as operating at any stage in the translation process, but is conventionally thought of as decision processes for reaching a unique interlingual representation during translation.

The CRL has been involved with Machine Translation practically since its beginning ten years ago. The XTRA system (English to Chinese) used a Semantic Definite Clause Grammar written in Prolog, and embodied a number of unique approaches to the treatment of conjunction, gapping, phrase and clause attachment, and the relaxation of semantic constraints. The ULTRA system translated text between five languages (Chinese, English, German, Japanese, Spanish) using an interlingual approach with bidirectional (symmetric) DCGs that served both as analyzers and generators. The system has a well-developed interface with multilingual capabilities for data entry, text editing, lexical entry, and data output. In addition to PANGLOSS, Machine Translation forms an important part of CRL research in other areas: a Japanese business letter generator (Bizlet), Intelligent Tutoring Systems (ITS), multilingual IR and information extraction (Tipster).

During the initial phase of the PANGLOSS project, the CRL's objectives were to develop tools for constructing lexical items and ontological entries automatically from on-line resources, to develop the initial Spanish analysis component, and, jointly with CMT and ISI, to establish the infrastructure for the three site project.

The second phase is a two-year program (now extended to three), and we are currently at the end of the first year of this phase. Building on the results of the initial phase, the second phase calls for the construction of a new analysis component, with a considerably broader base than the first year analysis system, and involving a wider variety of fail-soft techniques; incorporation of a large-scale ontology into the year-two system; use of the jointly-developed interlingua; and continued emphasis on automatic acquisition of lexical and semantic information.

At this point, a first version of the year-two analysis system (called the Panglyzer) has been completed and is undergoing expansion and refinement. The second year analysis system begins with a dictionary-based part-of-speech tagger, which operates with an average accuracy of about 93%. This is followed by a component which chunks the tagged text into small syntactic sections. Sections identified as proper nouns are sent to another component which classifies them as geographical, personal, or corporate names. Other segments are analyzed, semantic/lexical information is accessed, and all information is incorporated into a uniform representation by a constituent parser. These constituents are grouped into clausal structures. Within each clause, the



various syntactic and lexical readings will be rated, and the top candidates passed on to the interlingual and generation components.

With respect to automatic acquisition, several advances have occurred in the past year. CRL has provided the ontology with a sense-disambiguated hierarchy of nominal word senses drawn from Longman's Dictionary of Contemporary English. The hierarchy is derived from disambiguated genus terms and rooted in the semantic categories provided by Longmans. We have gathered from our bilingual dictionaries a large number of phrases correlated with translations for that phrase. Verb classes for Spanish verbs, used as the basis for the morphological analysis program that feeds the part-of-speech tagger, were provided by Collin's Spanish-English/English-Spanish Dictionary. The CRL has also contracted with EFE (Spanish newswire service) for a continuous line feed.

In the latest evaluation, the PANGLOSS project showed the greatest improvement in time (normalized human, post-editing time), though still slightly behind IBM and further behind the Dragon system (Lingstat). All systems were essentially equivalent in quality after post-editing, as might be expected –post-editing brings each raw translation up to the level of the human translator. For the first time this year, the PANGLOSS system was evaluated in a fully-automatic mode, providing a base-line for future measurements. All systems must be evaluated in a fully-automatic mode at the next evaluation, currently scheduled for January 1994. A key challenge for all the competitors is to beat their control system (SYSTRAN for CANDIDE, SpanAm for PANGLOSS) in quality of translation.

---

### **MT Research in Thailand and Linguistics and Knowledge Science Laboratory (LINKS)**

*Virach Sornlertlamvanich*

*[Edited from: AAMT Journal, no.3 May 1993, p.14-15]*

It was more than 40 years ago that researchers in Thailand showed an interest in putting a facility for processing the Thai language on a computer. First attempts were the development of a Thai word processor with draft character code assignment. Development was done on a mainframe computer, and thanks to great efforts by researchers from many fields, we now have no difficulty in creating and reading Thai documents at computer terminals.

The first project on the development of an MT system for English-Thai translation was conducted by a Thai researcher with technical support from Grenoble University in France. The ARIANE project started in 1981. The peculiarities of Thai caused a bottleneck in progress and many language-specific problems were left to be overcome. At the same time, research by linguists on the modern language suggested some systematic models of Thai. Though we have no exhaustive model of Thai at the present time many researchers express their enthusiasm in realizing a computer system with the capacity to parse and generate any Thai sentences.

As part of the Multi-lingual Machine Translation (MMT) system for Asian languages, the National Electronics and Computer Technology Center (NECTEC) and the Center of the International Cooperation for Computerization (CICC) agreed to cooperate in the research and development of an MT system for Thai. NECTEC and CICC launched the project in 1987 and expect to have a system which can translate among five languages Thai, Chinese, Malay, Indonesian and Japanese through an intermediary representation called an interlingua. We are assured that the interlingua translation method is plausible for the design of a system for multilingual MT. The project is scheduled to end at the beginning of 1995.

The project started with some anxiety about the research background for the Thai language and about the infrastructure of Thai researchers in the field. Computer scientists and linguists from universities joined discussions on the preparation of the linguistic data and adaptation to the parsing algorithm. The research and development plan has produced details for components of Thai sentence analysis, Thai sentence generation, Thai electronic dictionary of basic terms and technical terms, translation support, Thai input-output system, and integration system. With the research cooperation

of the King Mongkut Institute of Technology Thonburi (KMUTT) and the King Mongkut Institute of Technology Ladkrabang (KMUTL), the translation system has gradually improved quantitatively and qualitatively.

Thanks to the NECTEC-CICC MT research and development project many research institutes and universities are actively conducting research on natural language processing. Much research is focused on parsing of Thai sentences, such as CUPARSE, a Thai syntactic parser based on a dependency grammar approach, developed at Chulalongkorn University; a syntactic parser at AIT based on GPSG; etc. Others are researching specific aspects of the Thai language. Much work on Thai is paving the way for building a large-scale MT system, and in the very near future we hope that various MT systems able to handle Thai as source or target language will be unveiled.

The Linguistics and Knowledge Science Laboratory, known as LINKS, was founded as a national research laboratory in 1992. LINKS - a laboratory within NECTEC - is fully subsidized by the Thai government to conduct research in the fields of natural language processing and applied AI. In addition, LINKS is also part of the research strategy which will stimulate research activity in Thailand.

The goals of LINKS are as follows:

1. to study and investigate research on information processing, NLP, large-scale databases.
2. to research in NLP and AI for a feasibility study of system development
3. to join system development with the private sector so that they gain facilities, resources and funds for advanced research.

In MT development, LINKS is joining the NECTEC-CICC's MMT project, guiding the development enabling the Thai language system to connect with the other four languages through the interlingua. With great support from CICC the system can handle some specific Thai sentences which are carefully selected as the set of corpus sentences in system development. We are now making improvements and expect a usable MT system, which will be a tool for preparing information for data exchange among the participating countries.

In the development process, we are preparing a corpus to evaluate the accuracy of translation results. This corpus database together with the text database, word dictionary database and grammatical rule database are considered to be great resources for further research and development. We plan to do regular system evaluations by looking at the actual requirements for document translation.

MT technology is firmly established in recent years; computer scientists and linguists are working closely to resolve both computational and linguistic problems. Apart from general problems of computational linguistics, we find that it is not easy to analyze Thai texts into sentences - because we are not accustomed to writing texts sentence by sentence. It causes problems in the word and sentence recognition process. We do not have an effective algorithm for performing fully automatic word segmentation, and we are still facing difficulties in expressing syntactic structures. These are the fundamental issues that we must overcome in order to reach the next step of MT development.

Virach Sornlertlamvanich, Chief of Linguistics and Knowledge Science Laboratory (LINKS), National Electronics and Computer Technology Center (NECTEC), Ministry of Science, Technology and Environment, Thailand. Email: [virach@nwg.nectec.or.th](mailto:virach@nwg.nectec.or.th)

---

## **ATR demonstrates the Interpreting Telephone**

*[From: New Technology Japan April, 1993]*

A joint experiment in interpreting telephony was carried out on 1/28/93. The parties involved comprised Advanced Telecommunication Research's Interpreting Telephony Research Laboratories (Soraku-gun, Kyoto, Japan), Carnegie Mellon University (Pittsburgh, PA), Siemens (Munich, Germany), and Karlsruhe University (Karlsruhe, Germany). The experiment was the world's initial demonstration of trans-oceanic conversation utilizing interpreting telephone systems, linked by means of international circuits for real-time translation. ATR unveiled a system known as ASURA

(Advanced Speech Understanding & Rendering System of ATR), in which speech recognition, language translation and 'speech synthesis' are united. The speech recognition part uses an allophone model that can reflect the variation of phonemes (consonants and vowels) according to the other phonemes that precede or come after that. Utilizing the allophone model, the system can be fitted to any speaker utilizing a speech sample of around 10 words. It too employs grammatical knowledge to foresee the next phoneme in a sequence, and to eliminate ungrammatical word sequences. The system can handle around 1,500 words in speech recognition and language translation.

---

## **ATR Interpreting Telecommunications Research Laboratories**

*Yasuhiru Yamazaki (President, ATR)*

*[From: AAMT Journal, no.3 May 1993, p.11-13]*

**"Successful international demonstration of an automatic interpreting telephone between Japan, the United States and Germany" -- "A translating phone for overseas calls"** -- These are the headlines of recent major news reports on domestic and international television and newspapers on January 28 and 29 of 1993. For the first time in the world, this experiment linked Japan, the United States and Germany through international telecommunication lines and enabled the words of the speakers in each country to be conveyed in the language of the listeners, coming close to achieving a dream and winning the acclaim of ordinary people. With the great success of this international interpreting telephone experiment, the ATR Interpreting Telephony Research Laboratories have achieved a seven-year goal. As the next step, the ATR Interpreting Telecommunications Research Laboratories have been inaugurated in Kansai Science City. In this establishment, we are grateful for the cooperation and support received from Japan Key Technology Center and many funding corporations, with NTT at the forefront.

The new ATR Interpreting Telecommunications Research Laboratories have set "Basic research in advanced speech interpreting telecommunications" as the research target for the seven years from 1993 to the year 2000. The research activities at the new research laboratories will take over the pioneering work of its predecessor, the ATR Interpreting Telephony Research Laboratories, further develop each individual contributing technology, and then integrate them into a coherent system.

Current speech interpreting technology requires the person speaking to have adequate awareness of the interpreting system. For example, the speaker cannot be just anyone in the present stage. Also, the contents of the speech (i.e., the topic or subject domain), vocabulary, word segmentation, grammatical sentence, speaking speed and other such speech factors must be suitable for processing capabilities of the system. The ideal system would break out of these restraints and automatically and instantly convey the speaker's speech in the language of the listener, and would allow the speaker to speak naturally, without having any special consideration of the translating system. In particular, how wonderful it would be for Japanese people who have a keen awareness of the language handicap, to be able to speak freely with people from other countries, completely unaware of the language barrier. We could expect the range of activity of Japanese people in business and tourism to expand more and more...

As examples of linguistic phenomena that are difficult to translate, there are various expressions that depend on the conversational situation, context and fragmentary or ungrammatical idiomatic expressions. Computers are very weak at processing these types of expressions. It is necessary to break free of conventional methods of attacking these problems and seek a new approach. Thus our aim is to classify the problem into the following four technical areas and establish basic technologies for them.

### **(1) Recognition of spontaneous speech**

In ordinary conversation, transformations occur in the speaking speed, loudness of voice, frequency spectrum and other acoustic characteristics, and interjections such as "eh--" and "wu-n" and sounds

without particular meaning are inserted. For these reasons, the conventional static processing methods are unable to improve the speech recognition accuracy. To overcome this problem, a method for dynamically adapting and reconfiguring the prosodic model and the linguistic model according to the utterance situation must be established. Also, humans are capable of understanding the speech of unspecified speakers and can quickly adapt to the peculiarities in the speech of a new speaker. Conventional speech recognition employs the method of mapping the characteristics of the speaker onto those of a reference speaker. For this function to approach the human capability, it is of course necessary to establish a method to make mapping adaptive as well as to improve the mapping itself.

## **(2) Prosody processing and speech synthesis**

Human speech contains prosodic information such as accent and intonation and pan-linguistic information that implicitly expresses the intention of an utterance as well as linguistic information that has specific meaning and can be represented by text. Humans use both of these two types of information to correctly understand the meaning conveyed by an utterance. Accordingly, a method of extracting this pan-linguistic information in addition to the specific-meaning contents must be developed for an advanced speech interpreting telecommunications system. Research is also proceeding on speech synthesis techniques, including the construction of rules for controlling prosodic characteristics such as accent, intonation, tempo and rhythm of speech in order to produce natural and varied synthesized speech.

## **(3) Cooperative integrated technologies for distributed natural language processing and MT**

Natural dialogs contain fragmentary expressions that depend on assumptions concerning the utterance situation and the knowledge of the listener, and among them expressions with other intentions are inserted regularly, and what is more, inversions and corrections also appear. For this reason, translation on the basis of rules alone (rule-based translation) has its limits. One promising technique for overcoming these limits is the method in which the input speech is first analyzed into clause and phrase patterns, and the most appropriate example is extracted by finding the best match for these patterns in a previously compiled set of clause and phrase patterns and their corresponding translations. Thus, by analyzing the translation process into the partial problems of rule-based translation, example-based translation and translation based on content understanding, and then combining the solutions to all of the partial problems, we hope to establish basic technologies that can lead to the most accurate translation attainable.

## **(4) Integrated processing of speech and language**

An effective way to improve the understanding of natural spoken language is to raise the accuracy by sharing of information between the speech processing and the language processing, taking advantage of the mutual facilitating effect. For example, the speech recognition rate can be increased by predicting the following spoken expressions, using the linguistic function of interjections and adverbs, which seems to be often unimportant in the sense of having a clear and specific meaning. Conversely, intonation, accent and other prosodic information can be used to eliminate syntactic and semantic ambiguity during the language processing, making it possible to extract a particular intention. Turning attention to the fact that "utterance situation" is the point of contact between speech processing and language processing, we will study methods for linking the two types of processing. A large-scale database of actual spoken language data is extremely important to speech research, so we continue to produce such a database.

To aggressively tackle the four research areas described above, this research facility is organized into four laboratories and a planning section. The research staff includes many researchers from Europe and the United States as well as from the funding organizations, so in the laboratories we can hear Japanese, English, French and German words flying about. We have succeeded in creating the international environment needed for research on speech interpreting telecommunications. In this type of environment, we hope to pursue the research goal, or dream, and

proceed with the research in cooperation with domestic and international research organizations. We believe that in the coming 21st century, "global communication among different languages" will become a more and more realistic demand.

All of this laboratory's personnel will work together to make the advanced speech interpreting telecommunication technology of this research project useful in achieving such global communication. We hope for your continuing understanding and support to our laboratories into the future.

---

## SRI and Spoken Language Translation

*David Carter*

*[From SALT listserv]*

We are nearing the end of a one-year pilot project to develop a Spoken Language Translator (SLT). The participants are SRI International (Menlo Park and Cambridge), the Swedish Institute of Computer Science, and Telia Research (Stockholm).

The guiding strategy of our project is intelligent re-use of existing components: SRI Menlo Park's DECIPHER speech recognizer, SRI Cambridge's Core Language Engine, and Telia's ProPhon synthesizer. At the moment the system is configured to translate spoken utterances from the ATIS domain (air travel enquiries), with vocabulary about 1000 and perplexity about 60, into spoken Swedish.

We have achieved what we think is a very promising level of performance using a judicious mix of rule-based and statistical techniques:

Under office conditions, for speakers of American English, about 65% for sentences, within the domain, of 12 words or under, are both successfully recognized and fall within the system's analysis grammar coverage. By this I mean that the analysed string is either what the user said, or is close enough to it in meaning to be acceptable.

In a recent test of 200 unseen text sentences from the domain, 75% resulted in a translation being produced, and 90% of the translations produced were acceptable. We believe that within three to five years, we could be producing acceptable translations for 95-97% of correctly-recognized sentences; and an analysis of the current failures leads us to conclude that the human effort needed to achieve this is nearly all domain-independent (domain-dependent tuning being largely a matter of automatic statistical processing of corpora).

Future developments may include the introduction of other languages and/or speech access to databases, depending on what organizations become involved.

We presented a paper on the system at the recent Princeton human language technology conference, and demonstrated it at the Washington Spoken Language Technology Applications day in April. We will be demonstrating it again at the ACL conference in Columbus, Ohio, and we have a paper in September's Eurospeech conference in Berlin that describes the system more fully and attempts to substantiate the predictions made above. If anyone would like a copy of the Eurospeech in advance, please let me know your full postal address.

David Carter  
SRI International Cambridge Computer Science Research Centre  
23 Millers Yard  
Cambridge CB2 1RQ, U.K.

---

## English Constraint Grammar Parser Developed at the University of Helsinki

*[Extract from: LINGUIST List: Vol-4-429.]*

As of June 1, 1993, the English Constraint Grammar Parser ENGCG, developed at the Department of General Linguistics and the Research Unit for Computational Linguistics, University of Helsinki,

is released for non-commercial academic use. ENGCG is released in collaboration with Helsinki University Licensing Ltd.

The various parts of the system were written by Atro Voutilainen, Juha Heikkila, Kimmo Koskenniemi, Bart Jongejan, Pasi Tapanainen.

The system is shipped as a fully compiled run-time version for Sun SparcStations (2 or 10). (Depending on customer requirements, it may become available for other machines as well.)

ENGCG is based on the Constraint Grammar framework originally proposed by Fred Karlsson. The theoretical background as well as the English description is documented as a book to appear under the title: Karlsson, F., Voutilainen, A., Heikkila, J. & Anttila, A. (forthcoming): "Constraint Grammar: a Language-Independent System for Parsing Unrestricted Text". To be published by Mouton de Gruyter.

This offer concerns non-commercial academic research purposes. For further information: Atro Voutilainen, Dept. of General Linguistics, P.O. Box 4, University of Helsinki, FIN-00014 University of Helsinki, Finland. Email: avoutila@ling.Helsinki.FI; Fax: +358 0 191 3598

---

### **LDDS: Linguistic Data on Diskette Service**

#### **A new service by LINCUM Europa**

*[From the CORPORA electronic discussion group (corpora@hd.uib.no), June 15, 1993]*

LINCUM Europa offers a new service to linguistic researchers. Linguistic data of the following kind are published and distributed on diskette: word lists, data for comparative language studies, dictionaries, bibliographical surveys, encyclopedic collections, linguistic doctoral dissertations and M.A. theses of all linguistic disciplines are distributed on diskette by Linguistic Data on Diskette Service.

The advantage of this service is the faster access to hitherto unpublished material and facilitation of discussion. Data in all languages are accepted for publication on diskette.

Send information on your data on diskette to: Ulrich Lueders, LINCUM EUROPA, P.O. Box 1316, D-85703, Unterschleissheim/München, Germany. Fax: +49-89-314 89 09

---

### **Concordance software and text corpora from OUP**

*[From LINGUIST List: Vol-4-549.]*

**MicroConcord**, Mike Scott and Tim Johns. Manual written by Simon Murison Bowie. An MS-DOS concordance program for use with large texts of English or European languages. Oxford University Press.

**Corpus Collection A.** (MS-DOS) One million words of text from the British Newspaper, the Independent. The text consists of five 200,000 word mini corpora: home news, foreign news, business news, arts, and sport. Oxford University Press.

**Corpus collection B.** (MS-DOS) One million words of academic text divided into five 200,000 word mini corpora: physics and biology, medicine and psychology, history and law, the arts and religion, and philosophy. Oxford University Press.

For information in the US, contact Athelstan at 1 800 598 3880. Outside the US contact OUP at +44 865 267971.

---

### **IBM Announces Speech Recognition Breakthrough**

*[News release May 25, 1993]*

IBM announced yesterday at Comdex that it has successfully ported its most sophisticated speech recognition technology from high-end RISC-based processors to Intel\* 486 and Pentium\* chips, an innovation that puts its most powerful speech-recognition capabilities onto desktop PCs.

The IBM Speech Server Series (ISSS) provides the most accurate and sophisticated speech recognition capabilities available today. Designed by IBM Research and announced at Comdex last fall, ISSS is a multi-user system that runs on a RISC System/6000\* client/server platform. The system features up to a 32,000-word vocabulary and takes dictation at a throughput of more than 70 words per minute. The IBM Speech Server is ideal for healthcare professionals, journalists, attorneys, scientists and other business users.

This breakthrough will make it possible for PC users to access the same advanced capabilities of ISSS previously available only to those on Local Area Networks (LANs). More powerful Intel microprocessors, enhanced software algorithms and other hardware improvements enabled this advance, said IBM specialists.

### **Surprising the Industry**

Just six months ago, IBM speech-recognition specialists believed it would take several years for the capabilities of its RISC algorithms to be ported to PCs. But rapid advances, especially in speech-recognition software, have dramatically shortened the cycle.

"We may surprise many industry observers by being able to offer speech recognition technology on desktop PCs so soon after having announced it on the RISC System/6000 platform," said Elton Sherwin, manager, speech recognition strategy and market development, Personal Systems. "But we feel it's important to move ahead as quickly as possible with speech recognition, an exciting technology that promises to make PCs much easier and more intuitive to work with."

Mr. Sherwin added that this achievement represents a renewed effort by IBM to speed its product development cycle. IBM's new target is to go from initial research to product development in six months – and to market six months thereafter.

To further accelerate the development of speech-enabled technologies and applications, IBM is seeking to maximize the potential from this technology through industry alliances and partnerships. IBM is also exploring possible initiatives and development partners to port the technology for telephony applications, an emerging market with strong growth potential.

IBM's Speech Server can provide surprisingly high accuracy. The system can correctly distinguish between homophones, recognize the start of a sentence, and provide capitalization.

### **IBM's Speech Recognition Family**

IBM's existing speech recognition family includes three other products besides ISSS: IBM's Continuous Speech Series<sup>®</sup>, which lets software developers create speech-recognition applications that run on a RISC platform and 486 microprocessor; VoiceType 2<sup>®</sup>, for DOS applications such as word processing, databases and spreadsheets; and VoiceType Control for Windows, and commands and navigation in Windows<sup>™</sup> 3.1 IBM is also demonstrating ICSS and VoiceType 2 and VoiceType Control for Windows at Comdex.

ICSS was developed jointly by IBM Research and Carnegie Mellon University. VoiceType 2<sup>®</sup> and VoiceType Control for Windows were developed by Dragon Systems<sup>®</sup> Inc., of Newton, Mass., a leader in PC-based speech recognition systems.

IBM VoiceType Control for Windows will begin shipping in early June and will be widely available through computer retailers at a suggested retail price of \$129.00.

---

## **The Text Software Initiative**

*[Extract from: LINGUIST List, Sat 29 May 1993.]*

The widespread availability of large amounts of electronic text and linguistic data in recent years has dramatically increased the need for generally available, flexible text software. Commercial software for text analysis and manipulation covers only a fraction of research needs, and it is often expensive and hard to adapt or extend to fit a particular research problem. Software developed by individual

researchers and labs is often experimental and hard to get, hard to install, under-documented, and sometimes unreliable. Above all, most of this software is incompatible.

The Text Software Initiative (TSI) is committed to solving this problem by working to establish and publish guidelines and standards for the development of text software; and to promulgate and coordinate the development of free TSI-conformant software.

The scope of the TSI covers all areas of analysis and manipulation of all kinds of texts (written or spoken, mono-lingual or multi-lingual parallel, etc.), including markup of physical and logical text features, linguistic analysis and annotation, browsing and retrieval, statistical analysis, and other text-related tasks in research in computational linguistics, humanities computing, terminology and lexicography, speech, etc.

The TSI software development effort is distributed, that is, anyone can contribute on a voluntary basis. This means that tools will be developed according to the contributors' priorities; however, the TSI is ultimately working towards the development of a comprehensive text handling system.

Project Coordinators: Nancy Ide (Vassar College, Poughkeepsie, New York, USA; ide@cs.vassar.edu), and Jean Véronis, Université de Provence/CNRS, Aix-en-Provence, France; veronis@grtc.cnrs-mrs.fr)

---

## ALVEY Natural Language Tools (Release 4)

*[Edited extract from: SALT listserv]*

A fourth (and final) release of the Alvey Natural Language Tools (ANLT) is now available. The UK Alvey Programme originally funded three projects at the Universities of Cambridge, Edinburgh and Lancaster to provide tools for use in natural language processing research. The DTI and SERC has funded their continued support and enhancement. The tools, a MORPHOLOGICAL ANALYSER, PARSERS and a GRAMMAR and LEXICON, are usable individually as well as together (integrated by a GRAMMAR DEVELOPMENT ENVIRONMENT) forming a complete system for the morphological, syntactic and semantic analysis of a considerable subset of English.

The ANLT system is available by anonymous FTP from Cambridge University, Computer Laboratory. The files containing grammars, lexicons and source code are encrypted, however, reports describing the system, specimen licence agreement and other information is not. If after examining the documentation, you wish to purchase a licence for use of the system for research purposes, you should complete and sign the specimen agreement and return it together with a cheque for the amount specified in the agreement (currently 500 ECU -- 100 ECU upgrade -- or local currency equivalent) to:

Lynxvale WCIU Programs, 20 Trumpington St., Cambridge, CB2 1QA, UK. Fax: +223 332797

All of the software components are written in Common Lisp and have been tested in several implementations on a wide range of machines.

ANLT distribution arrangements and instructions, and a machine-readable specimen licence agreement are available in files on the FTP server ftp.cl.cam.ac.uk (128.232.0.56).

To fetch this information use anonymous FTP (login with user name anonymous, and password your e-mail address), go to the directory 'nltools', and fetch the following files:

*licence* a machine-readable specimen licence agreement

*instruct* instructions on how to FTP technical reports and the ANLT itself

---

## PUBLICATIONS ANNOUNCED AND RECEIVED

Journal of Quantitative Linguistics



[From LINGUIST List: Vol-4-545.]

A new journal on quantitative linguistics will be launched in 1994. Authors are invited to submit 4 copies of their article providing it is within the scope of the journal.

*Scope.* The Journal of Quantitative Linguistics is a new international forum for the publication and discussion of research on the quantitative characteristics of language and text in an exact mathematical form. This approach, which is of growing interest, opens up important and exciting theoretical perspectives, as well as solutions for a wide range of practical problems such as machine learning or statistical parsing, by introducing into linguistics the methods and models of advanced scientific disciplines such as the natural sciences, economics and psychology.

Specifically, JQL will publish papers on:

1) Observations and descriptions of all aspects of language and text phenomena, including the areas of psycholinguistics, sociolinguistics, dialectology, pragmatics, etc. as far as they use quantitative mathematical methods (probability theory, stochastic processes, differential and difference equations, fuzzy logics and set theory, function theory etc.), on all levels of linguistic analysis.

2) Applications of methods, models, or findings from quantitative linguistics to problems of natural language processing, machine translation, language teaching, documentation and information retrieval.

3) Methodological problems of linguistic measurement, model construction, sampling and test theory.

4) Epistemological issues such as explanation of language and text phenomena, contributions to theory construction, systems theory, philosophy of science.

*Audience.* The Journal of Quantitative Linguistics will be important reading for all researchers in the following disciplines who are interested in quantitative methods and observations: linguistics, mathematics, statistics, artificial intelligence, cognitive science, and stylistics.

The Journal is edited by Reinhard Koehler, Linguistische Datenverarbeitung, University of Trier, Germany and will initially have 3 issues a year. Contributions (4 copies) should be sent to: The Journal of Quantitative Linguistics, Editorial Office, Swets & Zeitlinger/SPS, P.O. Box 825, 2160 SZ Lisse, The Netherlands; or: The Journal of Quantitative Linguistics, Editorial Office, Swets & Zeitlinger/SPS, P.O. Box 613, Royersford, PA 19468, U.S.A. For further information please contact by e-mail: koehler@ldv01.uni-trier.de, or Scrivy@swets.nl

### **Terminology**

**Terminology** is a new journal that will be launched in 1994 under the editorship of Helmi Sonneveld (Amsterdam) and Kurt Loening (Columbus, OH). Juan Sager (Manchester) acts as Consultant and the Editors are assisted by an international and interdisciplinary team of associate and advisory editors.

**Terminology** will encompass terminology both in general (theory and practice) and in specialized fields (LSP). The focus will be on the discussion of (systematic) solutions not only of language problems encountered in translation, but also, for example, of (monolingual) problems of ambiguity, reference and developments in multidisciplinary communication. Particular attention will be given to new and developing subject areas such as knowledge representation and transfer, information technology tools, expert systems and terminological databases.

Each issue will contain in-depth articles, research reports, short notes, reviews of books and products and reports on the activities of national and specialized terminology organizations.

The Editors invite relevant contributions which may be submitted to: Helmi Sonneveld, Topterm CV, W.G.Plein 278, NL 1054 SE Amsterdam, The Netherlands. Fax: +31 20 6851591

**Terminology** is published in 2 issues per vol., appr. 400 pp. Subscription rate for volume 1 (1994): Hfl. 200/US\$ 120 (postage & handling included). Published by: John Benjamins Publishing Co., P.O.Box 75577, 1070 AN Amsterdam, The Netherlands. Fax: +31 20 6739773. (or: 821 Bethlehem Pike, Philadelphia, PA 19118, USA. Fax: +1 215 836 1204.)

## PUBLICATIONS RECEIVED

### *Journals*

**AAMT Journal:** Journal of the Asia-Pacific Association for Machine Translation, *no.2 February 1993*. p.1: MT Summit IV in Japan (Eiichiro Yoshikawa). -- p.2-4: Translation Fair '93 [see this issue]. -- p.5-6: Sharp Duet-Qt. -- p.6: CSK introduces ARGO J/E [see this issue]. -- p.7-8: ReadyPen. -- p.9-10: LogoVista E to J [see this issue]. -- p.11: What is a role of a translator in the era of machine translation? (Hiroko Motoyoshi). -- p.12-13: Oki Electric Industry Co.Ltd. -- p.13: MT System Evaluation Committee (Shoichi Yokoyama). -- p.14-15: The Toin Corporation (Takeshi Narumi). -- p.16-17: Machine Translation Evaluation Workshop/Showcase: San Diego report (Hirosato Nomura). -- p.17: Committee & Work Group logs. -- p.18-21: A short guide for MT techniques Part 6: Sentence generation techniques (Takenobu Tokunaga). -- p.22-23: Announcements: MT Summit IV [and] TMI-93. -- p.24: Products on market in Japan.; *no.3 May 1993*. p.1-5: Survey of MT usage in Japan (Takenori Makino) [see this issue]. -- p.5-10: Human translation and machine translation (Susumu Ando). -- p.11-13: ATR Interpreting Telecommunications Research Laboratories (Yasuhiru Yamazaki) [see this issue]. -- p.14-15: MT research in Thailand and Linguistics and Knowledge Science Laboratory (LINKS) (Virach Sornlertlamvanich) [see this issue]. -- p.16-17: Oki Electric Industry Co.Ltd. PENSÉE-GV [see this issue]. -- p.18-20: MATES/EK (MACHINE Translation Environment System/ English-Korean) (Gil-Chang Kim and Key-Sun Choi). -- p.21-26: A short guide for MT techniques Part 7: Dictionaries for machine translation (Hiroshi Suematsu). -- p.27: Announcement: MT Summit IV.

**Computational Linguistics**, *vol.19, no.1 (March 1993) - Special issue on Using Large Corpora I*: p.1-24: Introduction to the special issue (Kenneth W.Church and Robert L.Mercer). -- p. 23-60: Generalized probabilistic LR parsing of natural language (corpora) with unification-based grammars (Ted Briscoe and John Carroll). -- p. 61-74: Accurate methods for the statistics of surprise and coincidence (Ted Dunning). -- p.75-102: A program for aligning sentences in bilingual corpora (William A.Gale and Kenneth W.Church). -- p.103-120: Structural ambiguity and lexical relations (Donald Hindle and Mats Rooth). -- p.121-142: Text-translation alignment (Martin Kay and Martin Röscheisen). --p.143-177: Retrieving collocations from text: Xtract (Frank Smadja).

**Language Industry Monitor**, *no.15 (May-June 1993)*. p.1-5: TM/2: tip of the iceberg? [IBM activity in MT and translation aids.] -- p.5-8: ARPA's deep pockets [US government funding of language technology]. -- p.14: EAGLES stretching its wings; *no.16 (July-August 1993)*. p.1-6: Europe's linguistic superhighway [launch of CEC Language and Technology 2000 programme.] -- p.10-12: San Diego's other MT company [on MicroTac's Language Assistant series.] -- p.12-13: Eurolang: turning metal into gold.

**Language International**, *vol.5 no.3 (June 1993)*. p. 5-7: Translator's Workbench II (Sabine Bell). -- p.8-10: Where jokers make the rules (Ralph Dessau) [morphology in PC-Translator systems]. -- p.31-32: Translation debated in Århus [report of conference, January 1993, including papers on MT]

**Literary & Linguistic Computing**, *vol.8, no.1, 1993*. p.39-46: The large-scale production of syntactically analysed corpora (Roger Garside); *vol.8, no.2, 1993*. p. 85-90: An example-based system of writing English sentences for Japanese English users (Teiji Furugori and Akiko Takeda)

**Machine Translation**, *vol.7, no.3, 1992*. p. 135-193: The use of lexical semantics in interlingual machine translation (Bonnie J.Dorr). -- p.195-228: Book reviews; *vol.7, no.4, 1992/1993: Forum issue - Current research in machine translation*. p.231-246: Current research in machine translation (Harold L.Somers). -- p.247-253: Reaction to Somers: Current research in machine translation (Doug Arnold). -- p.253-258: Current MT research orientation/disorientation (Laurent R.Bourbeau). --

p.259-264: And after the second generation (John Hutchins). -- p.265-272: Current research in machine translation: a reply to Somers (Pierre Isabelle). -- p.273-280: State of the art and perspectives (Margaret King). -- p.281-284: Producers, users, customers: towards a differentiated evaluation of research in machine translation (Erich Steiner). -- p.285-292: Why MT systems are still not widely used (Keh-Yih Su and Jing-Shin Chang). -- p.293-302: What do we want from MT? (Muriel Vasconcellos). -- p.303-308: Responses to Current research in machine translation by Harold Somers (John S.White). -- p.309-314: Commentary on Somers' article (Yorick Wilks). -- p.315-322: A response to the comments (Harold L.Somers).

**Tribune des Industries de la Langue**, no.11-12, 1993. p.3-8: Renverser les mythes... (Stéphane Zadouniasky) [MT and MAT activity in Japan]. -- p. 17: STYLUS: système de TAO (André Abbou et Eléna Slavine) [system from St Petersburg: English-Russian, Russian-English, and French-Russian]. -- p.19-64: La convention "Ingénierie linguistique" 1993 [held by OFIL] - including: p. 22-23: An industrial view of linguistic engineering (T.Schneider); p.29-30: VERBMOBIL (Bernd Reuse); p. 53-55: Trados Translator's Workbench II (Gerhard Heyer); p.55-57: Les outils du traducteur au sein du service de traduction de la CCE (Hubert Paesemans); p.57-58: The Rinsche MT evaluation methodology (Adriane Rinsche); p.58-59: TECHDOC, a system for the automatic production of multilingual technical documents (Dietmar Rösner).

#### *Books*

**American Translators Association, Committee on Machine Translation: Report PC-based MT products.** [Arlington, VA: ATA], December 1992.

**Dorr, Bonnie Jean:** *Machine translation: a view from the lexicon.* Cambridge, Mass.: M.I.T.Press, 1993. xx, 432 pp. ISBN: 0-262-04138-3. Contents: Introduction (p.1-39); Syntactic processing (p.43-77); Morphological processing (p.79-92); The interlingual representation (p.95-162); Parameterization of the interlingua (p.163-192); Generation from the interlingua (p.193-231); Formalization of machine translation divergences (p.233-281); Translation examples (p.285-317); Current and future research: tense, aspect, and lexical acquisition (p.319-350); Conclusions (p.351-354); Appendices (p.355-393); Bibliography (p.395-422).

**Observatoire Wallon des Industries de la Langue:** *Des industries de la langue pour quoi faire? technologies, usages et marchés.* Liège: OWIL, 1993. 173 pp. [Survey and evaluation of systems and activities in MT, documentation management, office automation, and spoken language technology.]

**Rinsche, Adriane:** *Evaluationsverfahren für maschinelle Übersetzungssysteme: zur Methodik und experimentellen Praxis.* Bericht EUR 14766 DE (Kommission der Europäischen Gemeinschaften. Informationsmanagement). Brüssel, Luxemburg: EKGS Generaldirektion XIII, 1993. xii,211 pp. **Contents:** Maschinelle Übersetzungssysteme heute: eine Kurzübersicht (p.1-30); Evaluationsmethoden für MÜ-Systeme - Vorliegende Studien zur Evaluation von MÜ-Systemen (p.32-48); Beschreibung und Begründung des gewählten Evaluationsverfahrens (p.49-78); Emprischer Teil -Umfrage zum Einsatz von MÜ-Systemen (p.79-90); Studie I: Vergleichende Performanzevaluation von MÜ-Systemen für das Sprachpaar deutsch-englisch (p.91-99); Studie II: Evaluation linguistischer Performanz mit den Sprachpaaren französisch-englisch und deutsch-englisch (p.100-121); Verbesserung der Methode (p.122-127); Kritische Reflexion (p.128-130); 'Intelligente' Übersetzungssysteme - eine kritische Auseinandersetzung (p.131-154); Bibliographie (p.155-169); Anhang.

**Sonneveld, Helmi B. and Loening, Kurt L. (eds.):** *Terminology: applications in interdisciplinary communication.* Amsterdam, Philadelphia: John Benjamins Publ.Co., 1993. viii, 244 pp. ISBN: 90-272-2131-6; 1-55619-487-0.

## *Conference proceedings*

**31st Annual Meeting of the Association for Computational Linguistics.** *Proceedings of the Conference, 22-26 June 1993, Ohio State University, Columbus, Ohio, USA.* Contents include:

*The fourth Machine Translation Summit. MT Summit IV. Proceedings: International cooperation for global communication, July 20-22, 1993, Kobe, Japan.* Tokyo: AAMT for MT Summit Secretariat, 1993. [For details of contents see the report of the conference in this issue.]

*Proceedings of the Fifth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI '93. MT in the next generation, July 14-16, 1993, Kyoto Japan.* [For details of contents see the report of the conference in this issue.]

## *Reports*

**Linguistics Research Center, University of Texas.** *Quarterly report, March 1993.*

**Katholieke Universiteit Leuven, Centrum voor Computerlinguïstik.** *Working Papers in Natural Language Processing -- no. 11 (May 1993):* A computational interpretation of compounds (Lieve de Wachter & Jan Provoost); *no. 12 (May 1993):* Subsectional anaphora in DRT (Heleen Hoekstra).

**Ministry of Defence, Linguistic Services.** *Machine translation systems and their possible application within MOD Linguistic Services.* Short report. London: MoD, Feb. 1992. 32pp. [Full report not available outside MoD.]

---

*Items for inclusion in the 'Publications Received' section should be sent to the Editor-in-Chief at the address given on the front page. Attention is drawn to the resolution of the IAMT General Assembly, which asks all members to send copies of all their publications within one year of publication.*

---

## **Compendium of MT Systems**

In previous issues of MT News International, we have announced the intention of IAMT to compile a comprehensive listing of MT systems. Compilation of entries has begun based on existing sources of already available printed information. These will form the basis of entries which will be sent to relevant organizations, companies and research groups for their additions, amendments and ultimate approval. The hope is that the Compendium will be ready sometime during next year. It would greatly assist the editors in the compilation of this publication if all of those involved in the manufacture and development of MT systems could send the latest information about their systems to the Chief Editor, either directly or via the regional editors. Addresses are given on the front page. Please help us to make this an accurate and reliable source for information about MT, and thus to promote interest in MT worldwide.