

# MT News

## International

*Newsletter of the International Association for Machine Translation*

ISSN 0965-5476

Issue no.3, September 1992

---

### IN THIS ISSUE:

**Association news (JAMT/AAMT, EAMT, AMTA)**

**People on the move...**

**Conference reports**

**Systems and Projects**

**Users' Views**

**Publications**

**Conference Announcements**

**Forthcoming Events**

**Application and registration forms**

---

### *Editor-in-Chief:*

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

Fax: +44 (603) 58553; 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@crpculu.lu

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

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

---

## ASSOCIATION NEWS

### JAMT becomes AAMT

*Makoto Nagao*

The Japan Association for Machine Translation (JAMT) has agreed to extend its activities to Asian and Pacific areas after its one year's activity which included seven issues of JAMT Journal, a symposium (MT World 92) and an MT Workshop. The new association, established 17 June 1992, is to be called the Asian Pacific Association for Machine Translation (to be abbreviated as AAMT).

The association has been formed with the full support of MT and NLP researchers of several countries in Asian Pacific areas including China, Korea, Thai, Malaysia, Indonesia, Singapore, India, and Taiwan area. AAMT will be starting soon to distribute membership registration forms and the first issue of AAMT Journal throughout the Asian Pacific areas with the aim of attracting as many members as possible. The journal will retain the essential features of the past issues of JAMT Journal.

With the foundation of AAMT, the Japan association has brought its activities to an end

and has disbanded. All the board members of JAMT were nominated as board members of AAMT in order to maintain a continuity during the transitional period from JAMT to AAMT until the end of next March. During this transitional period new board members will be nominated. AAMT will endeavour to publish the AAMT Journal four times a year in English, and it will be organizing symposia, workshops and tutorials in various locations in the Asian Pacific area from time to time. We hope that those with interests in MT and related activities in this region of the world will support our programme.

---

## Articles of AAMT

[extracts]

"The association shall be called the Asia-Pacific Association for Machine Translation.

### *Purpose*

1. The purpose of AAMT is to contribute its every effort to the development of the machine translation systems in the world by establishing a common ground for the mutual exchange of any relevant information and the study and discussion of any subject common to the systems among manufacturers, users and research institutions interested in such systems, cooperating with any academic society or institution relating to such systems, and developing its activities in the international arena so as to realize the steady advancement and dispersal of such systems in the world.
2. AAMT shall be a noncommercial, nongovernmental association...
3. ...AAMT shall attempt to further develop its international activities in cooperation with the International Association for Machine Translation (IAMT) and other organizations affiliated to IAMT.

### *Business*

AAMT shall engage in the following business matters to achieve the purpose mentioned...

- (1) Holding of various meetings for research, lectures, discussions, training, educational visits, etc.
- (2) Publication of newsletters and any other form of literature, whether or not periodical.
- (3) Education of users for the machine translation systems, establishment of a training course for such users, and the conducting of tests for measuring technical skills in such systems.
- (4) Technical development of the machine translation systems.
- (5) Establishment of the standard and specifications regarding the machine translation systems, especially the preparation of any recommendation concerning the evaluation methods of such systems and the rules for the sentential constructions suitable for such systems.
- (6) Preparation of guidelines for the introduction and use of the machine translation systems.
- (7) Technical investigation and the collection and exchange of any information and literature.
- (8) Promotion of any academic or scientific investigation and study.
- (9) Establishment of any cooperative activity with any association, society, institution or organization of similar kind.
- (10) Establishment of any business other than mentioned in the preceding paragraphs necessary for achieving the purpose of AAMT."

The supreme body of AAMT will be the General Assembly of all members. This will appoint a Board of Directors (not less than nine and not more than thirty) and two Auditors. The Board of Directors, in turn, will elect a President and two Vice Presidents. Officers are appointed for two years and may be re-elected. The General Assembly is to meet at least once a year, under the

chairmanship of the President. The languages of AAMT will be English and Japanese, and its secretariat will be based, initially at least, in Japan with a Secretary-General appointed by the President.

Further details may be obtained from the present JAMT Secretariat (Mrs Megumi Okita), 305 Akasaka Chuo Mansion, 2-17 Akasaka 7-chome, Minato-ku, Tokyo 107, Japan

---

### **JAMT Journal and AAMT Journal**

With the disbanding of JAMT and the creation of the Asian Pacific Association for Machine Translation (AAMT), as reported above, the last issue of the Japanese-language JAMT Journal has also appeared. This was number 7 published in August 1992. It will be succeeded by two new publications: (1) an AAMT Newsletter, which will be in English and published as part of MT News International, and (2) an AAMT Journal, which will be published in both Japanese (for Japanese members) and English (for other AAMT members); it will be a quarterly (each issue 24 or 32 pages) and will contain items similar to those of the present JAMT Journal. Its first issue will appear in October 1992. For further information contact Professor Hirosato Nomura at Kyushu Institute of Technology (Iizuka, 820 Japan. Email: nomura@dumbo.ai.kyutech.ac.jp)

---

### **EAMT Committee meets during Coling**

The committee of EAMT held a meeting on Wednesday 29 July 1992 in a seminar room of the I.U.T., Nantes. The following were reported or agreed:

◁ the current EAMT membership numbers are: individual members 40; institutional members 7, non-profit making organisations 7.

◁ institutional members should receive five copies of the newsletter.

◁ telephone numbers, email addresses, etc. of all members of the EAMT committee should be distributed, so that members and others would have as many points of contact with the EAMT as possible.

◁ it was agreed that the Newsletter could be for sale to non-members (at a rate sufficient to make a profit), and the views of other Regional Associations and of IAMT Council should be sought.

◁ EAMT General Assembly should consider: Associate Members, with no voting rights but receiving the newsletter; lower rates for East European members; payment of fees in local currencies.

◁ the next EAMT General Assembly should be held at the European ACL next year.

The meeting received a report on the Machine Translation and Translation Theory Workshop held at Nantes, 22 July 1992, organized by Christa Hauenschild and Ulrike Schwall. The Workshop had been very successful. It had been attended by about 40 people and speakers had included Jan Albrecht (Heidelberg). It was proposed that another such meeting should be organized around the European ACL (Utrecht, April 1993).

Regarding future activities:

◁ EAMT is sponsoring two sessions at the Translating and the Computer 14 (Aslib Conference) being held 10-11 November 1992.

◁ Kirsten Falkedal (ISSCO, Geneva) is planning an MT evaluation meeting in Switzerland next year.

◁ the following future activities were proposed: an MT and the Lexicon Workshop to be held in Heidelberg next spring; a Conference/Workshop on MT and Corpora or MT and Machine Readable Dictionaries to be held in 1994.

◁ Geert Adriaens from Leuven had suggested that the next TMI and MT Summit V should be held in Brussels in 1995.

---

## **People on the Move...**

*Bernard (Bud) Scott*, founder of Logos Corporation and until recently its Chief Scientist, announces the formation of Parse International, a consulting firm specializing in natural language processing. Logos Corporation has signed on as Scott's first customer. Scott is a founding director of AMTA. His firm's address is P.O. Box 62, Ledgewood, NJ 07852, USA.

*Klaus Schubert* has left BSO/Language Technology BV, Utrecht, where he directed the MT project known as Distributed Language Translation and where subsequently he was principal consultant of the BSO subsidiary for software services in language technology. He has been appointed Professor of Computational Linguistics and Technical Translation at the Fachhochule Flensburg, as of 1 August 1992.

---

## **CONFERENCE REPORTS**

### **TMI in Montreal, 25-27 June 1992**

*John Hutchins*

The Fourth International Conference on Theoretical and Methodological Issues in Machine Translation took place in the Radisson Hotel in downtown Montreal. It was organized by the Canadian Workplace Automation Research Center (CWARC) - Centre canadien de recherche sur l'information du travail (CCRIT) - in collaboration with the Secretary of State of Canada, and was sponsored by the Canadian Department of Communications, the Canadian Department of the Secretary of State, the US Defense Advanced Projects Research Agency (DARPA), Bell Canada, the Association for Machine Translation in the Americas (AMTA), and the International Association for Machine Translation (IAMT).

In the reception preceding the conference on Wednesday 24 June the 150 participants were welcomed by the organizer Pierre Isabelle, who was delighted by the good response - whether due to increasing interest in MT, successful publicity, to the attractions of Montreal or to the program itself. He introduced Jacques Lyrette, assistant deputy minister at the Canadian Department of Communications, who spoke of the importance attached to MT and NLP by his department, saw CWARC as the spearhead of Communications Canada in this area, and mentioned the success of the MEDIATEX project (involving CWARC, IBM France and Société Grandjean) for real-time captioning of French TV programs. Klaire Tremblay, representing the Canadian Department of the Secretary of State (which is responsible for the Canadian Translation Bureau) spoke of her department's strong and long-standing commitment of MT, having supported TERMIUM, TAUM-METEO, the CWARC's translator's workstation and the testing of many commercial MT systems.

The conference proper was opened by Pierre Isabelle (CCRIT), who outlined the organization of the conference and introduced its general theme: the perceived polarization between the 'rationalist' methodology which has dominated MT research for more than two decades and the renewed interest in more 'empirical' methods. The former is characterised by an emphasis on rule-based approaches, founded on developments in theoretical linguistics and cognitive science. The latter emphasises the analysis of large text corpora using statistical, analogical and connectionist models. Although most participants questioned the labels

'rationalist' and 'empiricist' - all claiming to be empirical as opposed to the 'pure' theoreticians, and none seeing themselves as irrational - the two polarities were recognized in broad terms, and the organization of the presentations to reflect the tendencies was seen as a useful framework for present discussions of MT methodology.

The first papers were devoted essentially to lexical issues. Antonio Sanfilippo and Arturo Trujillo (both researchers at Cambridge University) described aspects of a lexicalist approach to translation, in which transfer is driven by bilingual lexical equivalences rather than by structural transfer rules - an approach dubbed the "shake and bake" method. They were followed by Laurence Danlos (Univ. de Paris), who argued that problems of aspect and modality could be handled easier by the rule-based approach of Eurotra than by 'purely' statistical approaches.

Connectionism was the focus of the next session. Ian McLean (UMIST) described an approach to example-based MT, in which a connectionist network would compute a distance measure between input text and items in a bilingual corpus of matching source and target translation examples. Akitoshi Okumura (NEC) and Masahiko Ishikawa (Matsushita) both presented 'hybrid' models for Japanese which combine traditional databases of syntactic and/or morphological rules and connectionist modules for 'learning' structural patterns from users' corrections and interactions, and for modifying and augmenting the rule base.

Lunch on the first day was followed by the first keynote speaker, Robert L. Mercer (IBM). He entitled his talk 'Rationalist MT: another cargo cult or just plain snake oil?', characterizing 'rationalist' MT researchers as people who are deceived by their own theories ("snake oil") and 'rationalist MT' as wishful thinking with no factual foundation ("cargo cult"). By contrast the 'empiricists' are not concerned with philosophical issues but with whether something works or not. The IBM group had set up a simple-minded translation model based on probabilistic methods which had been successful in speech recognition. The results are now familiar to MT researchers - with statistical methods alone they had achieved a level of accuracy approaching that of rule-based methods. Further improvement is expected with the addition of simple phonological and morphological information - a method claimed to be still empirical even if no longer purely statistical (a corruption of the previous IBM 'purity' regretted by a number of participants.) Mercer's prediction was that within five years all MT research would be empirical and the 'rationalist' approaches will be forgotten; and that empirical MT will be more successful than rationalist MT, just as empirical speech recognition has proved to be more successful than rationalist speech recognition, and the latter is now past history.

Appropriately, statistical methods were the focus of the next four papers. Michel Simard (CWARC) described an improved method of aligning bilingual text corpora by the using cognates (words in the two languages with identical orthographic properties, e.g. *generation* and *génération*, *error* and *erreur*). Peter Brown (IBM) re-interpreted the probabilistic IBM method in terms of the analysis-transfer-synthesis paradigm and described some further developments, including the use of traditional linguistic data. William Gale (AT&T) concentrated on the use of statistical methods for word sense disambiguation, using the Canadian Hansard corpus as database, and claiming to achieve up to 90% accuracy in sense discrimination. Finally Elliot Macklovitch (CWARC) examined the reasons for the failures of statistics-based syntactic tagging methods and discussed the incorporation of linguistic information to flag potential errors.

On the second day the focus moved on to example-based approaches. For Makoto Nagao (Kyoto Univ) they are means for overcoming those inadequacies of traditional grammar formalisms attributable to contextual limitations. Osamu Furuse (ATR) described the incorporation of an example-based method within the essentially transfer-based system for spoken language translation at the ATR Research Laboratories. Harold Somers (UMIST) outlined the design of a system for multilingual text generation via interaction with users, where input is matched with example target texts to ensure output quality. Daniel Jones gave further

information about the UMIST project for translating business letters and argued for its characterization as a 'pure' (non-hybrid) version of the example-based approach, i.e. one in which 'rationalist' (or rule-based) assumptions and representations play no part. Finally, Hiroshi Maruyama (IBM Research, Tokyo) presented a search algorithm for optimising pattern matching in an example-based approach.

At this point in the proceedings the response of the 'rationalists' began. It was opened by David Farwell (New Mexico State Univ), who defended the general plausibility of the hypothesis-testing methodology of rationalist approaches. Then Yorick Wilks gave the second keynote speech, which he had intriguingly entitled 'Stone soup and the French room'. He argued that there was no doubt that MT research was empirical (in the normal sense). The real question was whether MT is possible without explicit incorporation of linguistic data or 'knowledge'. The IBM statistical approach has been based on the premise that the system does not need to know any French (it is a 'French room' where as in Searle's Chinese room the operator does not know any Chinese to translate.) However, although it can translate, the results are poor - and the IBM researchers now propose improvements by adding extra ingredients: some morphology, some syntax, etc. (just as in the folktale when the 'soup' made of water and a stone is 'improved' by adding tasty vegetables and spices.) In other words, statistical methods alone have inherent limitations, and linguistic knowledge is needed - just as, Wilks conceded, the 'pure' linguistics rule-based approaches have limitations. The future of MT must inevitably lie in various kinds of 'hybrid' systems.

The following papers presented various 'rationalist' (rule-based) treatments of translation problems. Doug Arnold (Univ.Essex) discussed referential dependencies and their treatment in a LFG framework; Karin Haenelt (Darmstadt) described a discourse structure model as the basis for MT interface structure; Kurt Eberle (Univ Stuttgart) outlined an experimental bidirectional transfer system using a 'contextual resolver' module for dealing with pronouns and tense relations. The KANT system at Carnegie-Mellon University was the topic of a paper by Jaime Carbonell, who stressed the need to integrate a multiplicity of techniques from various MT paradigms (interlingua, transfer-based, statistical, etc.) if high-quality MT is to be achieved.

By the third day it was clear that most speakers really agreed with Yorick Wilks that 'hybrid' approaches were the future direction for MT research - and each of the final speakers presented some type of 'hybridisation': Hubert Lehmann (IBM Deutschland) argued that a combination of statistical and analytical methods are appropriate for discovering semantic compatibility relations; Keh Yih Su (Taiwan) described the corpus-based statistics-oriented Chinese-English system of the ArchTrans project; and Ralph Grishman (New York Univ) advocated the distributional analysis of analytically parsed tree structures for deriving transfer rules of an MT system.

Although hybridisation was by now the common flavour, the conference culminated in a final confrontation of the two extremes - with a 'disputation' in the medieval manner, where disputants argued for positions with which they did not agree. Graeme Hirst and Ronald Kaplan argued for the empiricists, and Geoffrey Sampson and Ken Church for the rationalists. The disputation was conducted on all sides with much wit and verve, and the adjudicator Martin Kay summarized the essential issues with his usual clarity and even-handedness. It is to be hoped that a transcript of the disputation will be made available for those who were not fortunate to be present. But at least the published proceedings will give some taste of the feast presented at this exceptionally well organized conference.

The proceedings can be ordered from: CCRIT, TMI-92, 1575 boul.Chomedey, Laval (Québec), Canada H7V 2X2. Email: [tmi@ccrit.doc.ca](mailto:tmi@ccrit.doc.ca)

---

## MT at COLING '92

*Steve Richardson*

Machine translation was represented by no fewer than 35 papers and project notes at the recent International Conference on Computational Linguistics (COLING), held in Nantes, France from 23-28 July.

The opening presentation of the conference was an invited lecture by Wolfgang Wahlster in which he detailed ambitious plans for a project sponsored by the German Federal Ministry of Research and Technology to create *Verbmobil*, a prototype speech-to-speech translation system, by the year 2001. ATR Interpreting Telephony Research Laboratories is also working on speech-to-speech translation for telephone interpretation, and hope to have a prototype that handles conference registrations ready for evaluation by the end of 1993. A paper by Tsuyoshi Morimoto provided specific information on the SL-TRANS2 system under development at ATR.

Another major project presented at the conference was *EuroLang*, jointly sponsored by a number of academic and industrial participants, which plans to produce a commercially-marketable second generation MT system by the end of 1994.

Papers presented by Keh-Yih Su and Jyun-Sheng Chang from National Tsing-Hua University in Taiwan documented their group's progress on using corpus-based statistical methods to perform and evaluate automatic translations. A paper by Shinichi Doi of NEC outlined the use of independent source and target language corpora to resolve ambiguity in MT.

Example-based methods for machine translation were widely discussed at the conference; they were the topic of papers by Hiroshi Nomiyama and Hideo Watanabe from IBM's Tokyo Research Laboratory, Hiroyuki Kaji of Hitachi's Systems Development Lab, Osamu Furuse at ATR, and Mitsugu Miura from NEC.

Lexical acquisition and maintenance, also topics of substantial general interest, were the subject of various MT-related papers, including those by David Farwell of CRL/NMSU, Teruko Mitamura from the Center for MT at CMU, and Brigitte Bläser from IBM Germany.

Machine-aided translation tools were described in papers by Satoshi Sato, now at JICST but formerly at Kyoto University, Eugenio Picchi at CNR in Pisa, and Huy Khanh Phan at GETA. Plans for the Translator's Workbench (TWB) II project were outlined which project a market-ready version by 1994. IBM's Translation Manager/2 was also demonstrated, although it was not in the official program.

Methods used in a variety of domain-specific translation systems were discussed in presentations by Masami Suzuki of ATR, Eric Nyberg at CMU, Bengt Sigurd at University of Lund, and K. Boesefeldt from ISSCO. A related paper on multi-lingual text generation in a specific domain was also presented by L.Iordanskaja of CoGenTex, Inc.

Architectures and methodologies used by specific systems were outlined in papers by Bonnie Dorr at University of Maryland (*interlingua*), Satoshi Kinoshita at Toshiba (*transfer*), Koichi Takeda at IBM's TRL (*KBMT*), Peter Whitelock, Chris Brew, and John Beaven of Sharp Laboratories (*shake-and-bake*), and Wilhelm Weisweber at Technical University of Berlin (*term-rewriting*). Finally, solutions for specific linguistic problems in MT, including article selection, delexical verbs, anaphora resolution, and aspect, were suggested in papers by Cornelia Zelinsky-Wibbelt (University of Saarland), Hideki Tanaka (NHK), Louisa Sadler (University of Essex), and Barbara Gawronska (University of Lund), respectively.

---

## **Second International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP).**

*Sofia Ananiadou*

The FG/NLP workshop was held in Manchester on the 30th and 31st of July 1992. It was organised by Sofia Ananiadou of the Centre for Computational Linguistics (University of Manchester Institute of Science and Technology) with the support of the ATR Interpreting Telephony Research Laboratories of Japan. It was attended by 60 participants.

The main topic of the workshop was to discuss the two current major trends in NLP, namely the newer stochastic approaches and the more traditional rule-based methods. Most of the papers were based on MT-related research. The papers given were as follows: Rule-based translation as constraint resolution (Louisa Sadler, Essex University); Grammar inference with lexicalized grammar (Yves Schabes, Pennsylvania University); Utility of massively parallel computing platform in natural language processing (Hiroaki Kitano, NEC Japan); Some successes in statistical MT (Vincent Della Pietra, IBM); The workstation substrate of the Pangloss project (Sergei Nirenburg, Carnegie-Mellon University); Linguistic knowledge acquisition from corpora (Jun-ichi Tsujii, Sofia Ananiadou, Iris Arad and Satoshi Sekine, CCL UMIST); Some rationales and methodologies for example-based approach (Makoto Nagao, Kyoto University, Japan); Transfer-driven machine translation (Osamu Furuse and Hitoshi Iida, ATR, Japan); Manual, automatic and machine assisted corpus annotation: the Lancaster experience (Geoffrey Leech, Lancaster University); Tagging an unfamiliar text with minimal human supervision (Eric Brill and Mitch Marcus, Pennsylvania University).

During the workshop members of ATR gave a demonstration of their Japanese-English MT system, which showed the effectiveness of their example-based approach. The workshop finished with a panel discussion summarizing the issues in the current debate in NLP research between traditional rule-based approaches and the newer stochastic and example-based approaches.

The proceedings of this FG/NLP workshop are available from Kate Stacey or Sofia Ananiadou, Centre for Computational Linguistics, UMIST, PO Box 88, Sackville Street, Manchester M60 1QD, UK. Email contact address (Dr.S.Ananiadou): [effie@ccl.umist.ac.uk](mailto:effie@ccl.umist.ac.uk).

---

**The Strategic Role of Evaluation in  
Natural Language Processing and Speech Technology.  
Edinburgh, 30 April - 2nd May, 1992.**

*Margaret King*

This three day, very intensive workshop, organised under the aegis of DANDI (the Esprit Working Group on Dialogue and Discourse), with additional support from the Human Communication Research Centre, Edinburgh, and ELSNET, the European Network for Language and Speech, brought together 15 people, who between them covered most areas of speech and language processing and who had, in addition, a good working knowledge of evaluation and assessment issues in the area.

The workshop was divided fairly evenly between advanced technical presentations and discussions, the latter continuing over breaks and into the evenings.

The aim of the presentations was to cover a wide range of different areas, in order to introduce ideas from one sector to another, to provoke discussion, and, if possible, to start the cross-fertilisation that is brought about by discussion with those who are outside one's immediate special domain but are close enough to it to understand it.

Thus, the scene was set with Henry Thompson setting the scene with a discussion of the strategic context of evaluation in the U.K. and Europe. He emphasized that the growing interest in evaluation issues meant that the whole community had to be prepared to take an active part in setting up evaluation methodologies and ensuring that they were adequate and fair.

The first group of presentations covered evaluation of spoken language systems, of



machine translation systems, of natural language generation systems, of information retrieval systems, and of speech recognition systems. Then particular topics were addressed by some of the speakers, such as the problem of evaluating semantic modules in the context of natural language understanding, the evaluation of speech synthesis techniques in a comprehension task and the special problem of multi-lingual speech input and output. Other speakers presented particular evaluation techniques, or focused on evaluations by developers of their own systems.

The overall consensus at the end of the meeting was that there was a great disparity between the sophistication of the technology available and the interest in its use across the different sub-areas of speech and language. The workshop had certainly fulfilled its aim of introducing ideas from one area to another, and enabling productive comparisons to be made. Perhaps even more valuable was the increased awareness in all of the participants of the status and history of evaluation and assessment attempts in other areas - a better idea, in other words, of where to go looking for good ideas, or for a critique of what seems like a good idea but hasn't been tried yet.

(This very brief summary has drawn extensively on Henry Thompson's introduction to the final report on the workshop.)

---

## **DARPA's Boston MT Party**

*Peter Brown and Eduard Hovy*

A new DARPA-sponsored MT initiative was begun during September 1991. Three projects were funded: the Candide system of IBM (New York), the Lingstat system of Dragon Systems Inc. (Boston), and the Pangloss system, jointly developed by the Center for Machine Translation (CMT) at Carnegie Mellon University, the Computing Research Laboratory (CRL) at New Mexico State University, and the Information Sciences Institute (ISI) of the University of Southern California.

Marking the close of the first of three years, and especially the results of the first formal evaluation, a group of 35 participating researchers, government personnel, and interested scientists met at Dragon Systems in Newton, Massachusetts on August 18 and 19, 1992.

Each of the three projects presented a system descriptions and an outline of future work. The Pangloss system addresses human-aided high-quality translation of newspaper texts about financial Mergers and Acquisitions from Spanish to English, using an interlingua. The first 10 months of development saw the assembly of the ULTRA parser from CRL, the Penman generator from ISI, and several user assistance tools embedded into the Translator's Workstation built at CMT.

The Lingstat system addresses human-aided browsing-quality translation of similar texts from Japanese to English, using a combination of statistical symbolic and methods. Lingstat was demonstrated on PCs, incorporating a Japanese tokenizer and a variety of user tools and dictionary resources.

The Candide system primarily addresses unaided MT, though it offers a text editor for post-editing as well. The system seeks to provide translations from French to English in any domain, using statistical techniques guided by some linguistic insights to collect the basic bi-language correspondences from 3 million sentences of the bilingual Canadian Parliamentary records.

A high point of the meeting was the discussion of the evaluation of the three systems and several others, namely SYSTRAN in Japanese, Spanish, and French, and PAHO's SpanAm for Spanish to English. In July of 1992, DARPA sponsored an objective multi-site evaluation of machine-aided and machine-alone translation systems on previously unseen texts. This event

infused the machine translation community with an evaluation methodology similar to that used in the DARPA speech recognition community.

The evaluation was conducted by John White and Teri O'Connell of PRC, McLean, Virginia, and Lynn Carlson of the US Government. Eighteen documents were translated into English by each system. The eighteen documents included six documents which were originally written in a source language and twelve documents which were originally written in English and then translated into different source languages. Source language translations of the twelve English documents were known as master passages. All documents were abstracts between 300 and 500 words in length from news articles on the subject of mergers and acquisitions. Documents were translated into English by translators working without the aid of a system, translators working with the aid of a system, and by systems working without the aid of a translator. The translators selected for the human-alone and human-assisted tasks were of intermediate skill level. All translations were judged for quality by a panel of evaluators which graded each sentence on a 12-point scale. Sentences were assessed penalty points for syntactic, semantic, lexical, stylistic, and orthographic errors. The evaluators were all senior translators fluent in the foreign language and native speakers of English.

All translations of master passages were also judged for comprehensibility. For each such passage, a set of multiple choice questions was constructed. Monolingual subjects then took comprehension tests on the original English passages and all the various translations back into English of translations of these passages.

In the machine-aided phase of the evaluation, the time it took translators to translate documents into English was measured. Times were measured for translators working with the aid of a system and for translators working without system aid. All translation times were normalized by the times it took the translators to translate a set of normalization passages without the aid of a system. In some cases a speedup in translation efficiency of as much as 35 percent was observed.

Participants in the August meeting universally felt that the evaluation was an extremely valuable experience, helping them to focus on aspects of their systems that needed improvement. It was agreed to hold another similar evaluation next spring. Consideration is being given to the possibility of opening this evaluation to other systems. System developers wishing to participate in future evaluations should contact IBM's Peter Brown (email: [pbrown@watson.ibm.com](mailto:pbrown@watson.ibm.com), phone: 914-784-6514).

In addition to the system status reports and evaluation results, Lingstat, Candide, and Pangloss were demonstrated, as well as Dragon's speech recognition system DragonDictate in several languages. The vexing question of acquiring and sharing textual, grammatical, and lexical resources was discussed. Frank Smadja from Columbia University described automated resource acquisition using statistical methods.

Finally, several invited participants, including Ralph Grishman from NYU, Lynette Hirschman from MIT, and Muriel Vasconcellos from PAHO gave their impressions of the program and the evaluation.

The meeting was generally deemed interesting, lively, and informative, and can easily be broadened into a workshop with more participation from non-DARPA researchers in the future. For more information, please contact Dr. Eduard Hovy, ISI (email: [hovy@isi.edu](mailto:hovy@isi.edu), phone: 310-822-1511).

---

## **SYSTEMS and PROJECTS**

**Eurotra**

*Doug Arnold*

This note gives an idea of what is happening, and is likely to happen in the Eurotra programme, and its surrounding initiatives.

### *1. Pre-industrial Eurotra*

Eurotra is a European Community (EC) funded Research and Development programme, with three aims:

(i) The development of a 'pre-industrial' prototype Machine Translation system 'of advanced design' (i.e. transfer based, and incorporating the technical and scientific innovations of the 70's and 80's), for the official EC languages (Danish, Dutch, English, French, German, Greek, Italian, Portuguese, and Spanish). Today this would probably be called a 'feasibility of concept demonstrator'.

(ii) Research on topics in computational linguistics (CL), MT and Linguistics which broadly relate to (i);

(iii) The development of infra-structure to support computational linguistics (CL) and natural language processing (NLP), and to support computationally applied language studies throughout the member states of the EC.

The official launch of the Eurotra programme was in November 1982, though a certain amount of work had been done before this. The 'pre-industrial' phase of the programme came to an end in December 1990. At its largest, it involved almost 200 staff, from 17 academic institutions throughout Europe, working on each of the above 9 languages (72 language pairs), some of which were essentially virgin territory for computational description. The total expenditure over this period was in the order of 40 MECU (roughly 40 million dollars); by comparison, the EC contribution to the ESPRIT initiative in information technology in 1987-92 was about 1,600 million ECU, which was itself about one third of the total EC contribution to R&D in this period.

Eurotra is certainly one of the largest, and most ambitious programmes of NLP R&D ever undertaken. From an infrastructural point of view, the programme has been a success, and a large and impressive body of linguistic research has been produced. But the developmental achievements are less impressive: although interesting scientific prototype implementations exist for most components of the system and, as intended, there is a 'pre-industrial' prototype which may be said to have demonstrated the feasibility of the concept, the achievement does not bear comparison with actual industrial/commercial systems in terms of speed, size of dictionaries, breadth of coverage, robustness, etc.

The 'pre-industrial' phase of Eurotra ended in December 1990, and was succeeded by the Transitional Programme for Eurotra. This phase, and the Programme as a whole comes to an end in December 1992. The transition phase is unlike the original Eurotra programme in making explicit provision for a number of separate and complementary projects alongside the core R&D activity, and in involving non-University research groups. Thus it is a response to the need for diversification in MT and CL research. From January 1993, research in this area will fall under the CEC's programme of *Linguistic Research and Engineering*, which has far wider aims, and a different organisational structure, and mode of financing.

### *2. The Transition programme (1991-1992)*

In this current phase the aims of the programme concern mainly the preparation of an operational Eurotra prototype system and the reusability of resources (lexical, terminological and grammatical). This broadening of aims implements one of the recommendations of an evaluation of previous phases of Eurotra (the Danzin report).

In the final year of the previous phase (1990), a number of 'feasibility and design' studies were undertaken. These dealt with the development of tools and methods for the reuse of lexical

resources in computer applications, assessed the strengths and weaknesses of the current prototypes with respect to the state of the art in CL and NLP and proposed an improved framework for future work. A number of high level requirements were placed on the redesign of the formalism, including a requirement that the design had to be totally 'mainstream' and capable of extension for the addition of new phenomena and capabilities. One study (ET-6/1) developed specifications for the new 'ET-6 Formalism', a second study (ET-6/2) led to specifications of a user and grammar development environment, and a third (ET-6/3) dealt with issues of low-level text encoding and handling (including some morphological analysis).

In the Transition Phase itself the following four activities are being pursued:

(a) A continuation of R&D work within the current framework, especially contrastive research on linguistic topics, aimed both at improving the current prototype and at providing a solid basis for work in the future (beyond the useful life of the current prototype);

(b) Implementation of an enhanced system for development and research, along the lines specified in the ET-6 studies (formalism, development environment, etc.);

(c) Research on a number of key topics, where the costs are shared equally between the CEC and industry (or funding authorities in member states);

(d) Training, carried out mainly in the participating centres.

As regards (a) research is being carried on monolingual topics with the ultimate aim of providing better treatments of particular phenomena in individual languages (for example, work on English, at Essex University, focuses on the treatment of idioms and modifiers/adjuncts). In the area of contrastive research, the aim is to provide linguistic analyses leading to representations which minimize differences between languages. This activity continues the model of the time/tense research in the pre-industrial phase, when a language-independent representation of time-reference was developed to replace language-specific representations of morpho-syntactic tenses. The contrastive topics receiving particular attention include: determination, negation and quantificational phenomena (which all involve notions of scope); compounding; and aspectual phenomena; sentential complementation; terminology; argument/modifier structure and 'interlevel syntax' (mappings between levels within analysis and synthesis). Research into transfer methods, especially 'reversible transfer' and 'relayed transfer' is also being carried out.

Work on (b) is being undertaken by a consortium of European software companies, essentially producing an industrial implementation based on the outcome of the ET-6 contracts.

As regards (c), the original call for proposals suggested a wide range of possibilities: research on translation theory, general topics of CL (morphology, interaction of lexicon/dictionary and grammar), knowledge representation and terminology, and research on the application of subparts of the existing system. In the event, six projects have been funded, all of which will take

ET-6/1 as their reference model. These are:

i) *Collocations and the lexicalization of semantic operations*. This concerns the extremely important problem of dealing with collocational restrictions (e.g. 'rancid butter' vs 'sour milk') in MT.

ii) *Terminology*. The objectives of this project include the definition of the internal representation of terminological definitions and their use in analysis and generation, the (semi-) automatic parsing of definitions and the use of the output of such parsing in analysis and generation.

iii) *Knowledge Bases*. This project involves an assessment of the feasibility and effectiveness of the (semi-)automatic parsing of dictionary definitions (from the COBUILD dictionary) as a form of knowledge acquisition for ET-6, with wider relevance for other natural language systems.

iv) *Implementation of Probabilistic and Corpus-based methods*. This project will investigate the possibility of adding a probabilistic component to the ET-6 architecture.

v) *Reusability of Grammars*. The research on the migration of grammars to the new (ET-6) formalism will also bear on the general issue of reusability of linguistic resources, which is likely to be increasingly important in the future.

vi) *Formal Semantics for Discourse*. This project is intended to take the first steps in a longer term research programme in the field of discourse and computational semantics.

The selection of topics for funding shows the importance currently given to lexical issues in current NLP, to the remediation of deficiencies in existing Eurotra dictionaries, and to the complementation of the linguistic and rule-based approach adopted in the pre-industrial phase by supporting investigations of how other kinds of knowledge can be added to a linguistic, rule-based system.

### 3. Future Prospects

CEC initiated R&D is organized into a series of framework programmes, each divided into 'Action Lines'. The Third Framework Programme (1991-4) includes seven Action Lines under the TELEMATICS program (this is comparable to, but smaller than, the ESPRIT action line, which relates to Information Technology). One of these action lines, headed 'LRE' (Linguistic Research and Engineering, area 6), is intended to support work on NLP of strategic importance, developing basic linguistic technology, "...with a view to overcoming limitations and inefficiencies brought about by the use of different languages within the Community" (LRE Call for Proposals), i.e. overcoming the (multilingual) language barrier.

The total budget for LRE is about ECU 22 million, with most of the research being organized on a 'shared cost' basis (though the plans also allow for some research that is wholly CEC funded, via grants, and study and service contracts, where this is appropriate, e.g. as with III below).

Work under LRE is grouped into five main headings. Calls for proposals under three of these headings (I, II, and IV) were published in November 1991, and the first set of projects are in the final stages of negotiation. Other calls are anticipated in 1992 and 1993.

The headings are as follows:

I. *Research of General Interest*: The main emphasis is research on: ways of increasing the interlinguality of linguistic representations of text/discourse; the use of domain specific knowledge (e.g. terminological, 'real world' specialist, and 'heuristic' knowledge); interfacing NLP and speech technology; and the use of 'advanced computational technologies' (to ensure that NLP R&D keeps pace with progress in computing).

II. *Common Tools and Resources*: The aim is to develop 'generic' software tools, grammars, dictionaries, terminological collections, and text corpora, which can be reused for a variety of applications and purposes (and hence are in some way 'theory neutral'). Typical software tools would be integrated testing and development environments, tools for dictionary construction and use, corpus analysis tools, 'workbenches'.

III. *Linguistic Standards*: The definition of commonly agreed data encoding schemes and formats for linguistic resources (e.g. dictionaries, grammars, corpora). It is clearly related to the previous heading. The kinds of actions foreseen include the setting up of European 'expert groups' to begin working towards the formulation of such standards, experimentation with existing candidate standards, and support for similar national and international initiatives.

IV. *Applications*: The aim is to support pilot and demonstrator projects in areas such as: MT; automatic document abstracting and indexing; aids for mono- and multilingual document generation, storage and retrieval; human-computer interaction; construction of Knowledge Bases from Natural Language Text; and Computer Aided Instruction. Such projects demonstrate the

feasibility, and manner of application of work under other headings, measure progress, and provide a way of testing results.

V. *Supporting Actions*: This covers training, initiatives to raise awareness, gather, synthesize, and disseminate information about NLP, with special emphasis placed on examining the economic and social impact of the technology, and legal problems that act as barriers to the emergence of new products and services.

#### 4. Literature

A good general discussion of the Eurotra project (i.e. Eurotra-1) can be found in Raw et al. (1988). For independent evaluations of 'pre-industrial' Eurotra see Danzin (1990) and Pannenberg (1987). The primary source of detailed discussion of Eurotra linguistics is the *Reference Manual* (the current, and more or less final version is 7.0, earlier versions are of mainly historical interest). More discussion of the linguistics can be found in Copeland et al. (1991), and in the special issues of the journal *Machine Translation* dedicated to Eurotra (vols 6 (2), and 6 (3), 1991). A description of the 'new' ET-6 formalism can be found in Pulman (1991), the results of the feasibility study on the development environment can be found in Schütz (1991), and Heid and McNaught (1991) include a discussion of the reusability of lexical resources.

Copeland et al is the first of a series of volumes in *Studies in Machine Translation and Natural Language Processing*, which is intended to embody the research results of the pre-industrial phase of Eurotra and the Transition Programme, in so far as they are not reflected in the Reference Manual. A second volume describing the Eurotra Formal Specifications has also appeared [see issue 2 of *MT News International*.] Other volumes in this series will consider: Assessment of Computational Linguistic Formalisms, Morphology, Interlevel Processing, Support Verbs, the Argument Structure of Nouns, and Preference Mechanisms (i.e. methods for choosing between competing analyses), Lexical Issues in MT, Reusability of Lexical Resources, Machine translation and discourse structure. [For information on ordering these, contact Erwin Valentini, DGXIII, Batiment Jean Monnet, Plateau Kirchberg, BP1907, L-2920, Luxembourg.]

C. Copeland, J. Durand, S. Krauwer, B. Maegaard, eds. (1991): *The Eurotra linguistic specifications*. Luxembourg: Office for Publications of the CEC, 1991 (Studies in Machine Translation and Natural Language Processing, 1.)

A. Danzin (1990): *Eurotra Programme Assessment Report*. Luxembourg: CEC, DG XIII, March 1990.

*Eurotra Reference Manual, 7.0*. Luxembourg: CEC, DG XIII, 1991.

Ulrich Heid, John McNaught, eds. (1991): *Eurotra-7 Study: feasibility and project definitions study on the reusability of lexical and terminological resources in computerized applications*. Luxembourg: CEC, DG XIII, August 1991.

A.E. Pannenberg, chairman (1987): *Eurotra Assessment Panel Final Report*. Luxembourg: CEC, DG XIII, October 1987.

S.G.Pulman, ed. (1991): *ET6/1: Rule formalism and virtual machine design study*. Cambridge: SRI International, 1991.

A. Raw, B. Vandecapelle, F. Van Eynde (1988): 'Eurotra: An Overview' *Interface: Journal of Applied Linguistics*, 3 (1), 1988, 5-32.

Jörg Schütz, ed. (1991): *Feasibility study for a Eurotra-II system: a feasibility study on the software environment with the framework of the Eurotra programme*. Luxembourg: CEC, DG XIII, 1991.

---

**EUROLANG**

*Eurolang Managing Board*

The main aim of this project is to develop a second generation machine-aided translation system (MT) in three years for the following language pairs: French↔English, German↔English, French↔German, Spanish↔English, Italian↔English.

By the year 1995, this portable industrial system should provide serious competition for Japanese products which are developed with considerable financial backing from the Japanese government and industrialists.

This realistic project will build upon the investments made by industrialists and the European public authorities over the last twenty years.

The SITE and SIEMENS NIXDORF groups, which control both European second generation industrial technologies (METAL and ARIANE), have decided to pool their technical, financial and human resources in order to develop a range of innovative products which reflect the industrial state-of-the-art as regards both the software and the hardware. A number of prestigious industrialists (KRUPP, MATRA, MARCONI SPACE, CAP GEMINI, RANK XEROX, etc.) and some of the most highly skilled European academics have also agreed to share their knowhow and resources.

The system is aimed at the market for quality translation with post-editing in the following fields: technical and commercial documentation, e-mail, multilingual querying of information systems, multilingual EDI and personal MT.

The EUROLANG project has just been launched within the EUREKA structure and will be completed at the end of 1994, at which time commercial releases of the MT products will be marketed. At the end of 1992 a prototype on a workstation will demonstrate the feasibility of future developments as well as the ergonomics and the productivity of the products.

Analysts have estimated the value of the world translation market at 12 billion dollars (250 000 000 pages). There is a great divergence between the demand and the production capacity of the traditional translation department. Only 5 to 10% of the demand can currently be met for a number of well-known reasons (cost, time factors, shortage of qualified translators). The products proposed by the consortium will target the "high quality, low cost" market, with an average return on investment of 100 FF per translated page.

The SITE group, the European leader in multilingual multimedia documentary engineering, subsidiary of the CORAREVILLON group, and SIEMENS NIXDORF are joint prime contractors for the project. The SIEMENS NIXDORF and SITE translation departments will be used to evaluate the products under development.

The system architecture is based on a toolbox. This will ensure reusability of the components and will permit additional applications in information retrieval, document generation and management.

This system should give the European business community the command of multilingual technical and commercial communication which is crucial for European integration and should surpass the Japanese competition in this strategic field.

---

### **Caterpillar Project at CMU**

During the TMI-92 conference in Montreal, Jaime Carbonell gave some details of the contract signed in May 1992 between Caterpillar, the world's largest manufacturer of earth-moving equipment, and the Center for Machine Translation at Carnegie-Mellon University for the development of a fully automatic translation system. The five-year multimillion dollar contract had been concluded after an extensive evaluation by Caterpillar since the 'proof-of-concept'

demonstration by the CMU team in June 1991. The system will be based on the research conducted at CMU on the knowledge-based interlingua system KANT. Input texts will be in controlled language (Caterpillar Technical English), and the documents for translation will be user and maintenance manuals - amounting to over 100 million words a year. For this restricted domain the system may eventually have a lexicon of some 60,000 word and phrase entries. The aim is a multilingual system translating at a speed of 100,000 words an hour from English into French, German, Japanese (and other languages later), with output requiring only minimal post-editing.

---

### **PAHO's Systems on PCs**

The Pan-American Health Organization (PAHO) is planning to introduce the PC version of ENGSPAN (tm) at the MT Showcase in San Diego. The system has been rewritten in the C language and runs under MS-DOS. The beta version is currently installed on PAHO's Novell network; the speed of translation is approximately 300 words per minute on a 386/33 processor, and 430 wpm on a 486/33. Optimization has barely begun, we are told.

The system includes a WordPerfect interface, dictionary browse and update utilities, dictionary and job accounting logs, and a Spanish spelling checker.

The program requires 500K to run. The conversion of the PL/1 code to C was begun in January 1992 and completed in six months. According to Marjorie Leon at PAHO, the PC version of SPANAM (tm) is on schedule and should be ready before the end of 1992.

---

### **CompuServe Investigates Machine Translation**

*Cole Harrison*

CompuServe, Inc. is investigating the possible use of machine translation in its on-line information service. With over one million subscribers, CompuServe Information Service is one of the largest on-line services in the world. It offers such information as newswires, financial quotations, weather reports, games, e-mail, on-line conferences in about 200 areas, and reference materials, with particular strengths in PC hardware and software support. Users typically connect through a PC and modem, and pay a flat rate per connect hour or per month. CompuServe's network has nodes in hundreds of cities, so the user does not pay long-distance telephone charges.

Most usage is currently in the U.S., Canada, and the U.K., but CompuServe is now making a major push to sign up users in Germany, with plans to follow up in the Netherlands, Belgium, and France. With increasing numbers of subscribers speaking different languages, CompuServe hopes to smooth communications among them by use of machine translation.

CompuServe's basic goal for MT is to provide draft-quality translation directly to end users. For example, it is hoped that German and French users can access on-line English information and download it in their own language. We suspect that there is a market for low-cost translations, even if the quality is less than ideal.

MT systems work most successfully with well-formed input texts and with a restricted domain of subject matter. CompuServe texts are subdivided into many small domains, but some of them are not very well-formed. The hope is that by providing special lexicons for many narrow domains and using the existing ability of our information service to sort the texts into the appropriate domains, that we will be able to produce translations which are at least somewhat understandable.

Our project will be in the evaluation stage for about six months, during which time we will survey existing English↔German systems using our texts and evaluate output quality, the



degree of improvability within a given domain, and computational performance. If the results of the evaluation are successful we hope to offer our subscribers an initial version of a machine translation service sometime in 1993. Work with other languages will also follow if the work with English↔German is successful.

We would be interested in hearing from researchers and users who have worked on the problem of providing draft-quality translation to end users and who have experiences to share. We are also currently searching for a project leader for this effort, who should have skills in system analysis and project management; knowledge of machine translation or computational linguistics; and be fluent in German. Please direct correspondence on either subject to: Cole Harrison, Director of Development, CompuServe Data Technologies, 1000 Massachusetts Avenue, Cambridge, MA 02138 USA. Tel. 617-661-9440; fax. 617-661-7858; e-mail: c.harrison@csi.compuserve.com

---

### **Intergraph's DP/Translator**

Intergraph Corporation has announced the sale to AT&T's Document Development Organization of Winston-Salem, North Carolina, of all eight language pairs available as part of DP/Translator, the company's integrated machine translation and desktop publishing system. These language pairs are English to French, German, Spanish, Italian and Portuguese, French to English, German to English, and Spanish to English. The Document Development Organization is responsible for the creation and translation of large volumes of AT&T internal as well as end-user documentation.

Stephen Kerce, manager of Intergraph's Natural Language group, also informs us that some of Intergraph's new installations include ZML Software Systems, Inc., in Louisville, Kentucky (English-French), Paul H. Brink International in Minneapolis, Minnesota (English to French, German, Italian, Portuguese, and Spanish), Occidental Exploration and Production Co. in Bakersfield, California (Spanish-English), and Matra Transit in Chicago, Illinois (French-English). Details of four more sales will be announced shortly, Dr. Kerce said.

Finally, we note that DP/Translator will be exhibited in the following forums in the near future: Seybold, San Francisco, 23-25 September; MT Showcase, San Diego, 2-3 November; and ATA, 4-8 November, also in San Diego.

---

### **LOGOS Announces New Release of MT Software**

Logos Corporation has introduced Release 7.0 of Logos multilingual document translation software for UNIX workstations. The software runs on either a Sun SPARCstation or an IBM RS/6000 and offers English and German as source languages and English, French, German, Spanish and Italian as target languages.

"Now that the Logos system has been ported to UNIX workstations and integrated with document publishing, we have accelerated our development activities," said William Hohenstein, Chairman. Completion of the systems integration development phase enabled Logos to obtain additional funding, which will be directed toward product development.

As mentioned in the announcement, Logos translation software provides document publishers and professional translators with complete, easy-to-use, interactive support systems for dictionary and semantic rule updating, revision processing, and post-editing. The company further states that texts are translated by the system at a speed of 15,000 words per hour. Logos is fully integrated with Interleaf and Frame electronic publishing software and WordPerfect 5.1, and is compatible with most other word processing software. Other technologies such as Xerox Imaging Systems' ScanWorX can be used as document input devices.

For more information about the Logos system, contact Deborah H. Bombaci, Manager of Market Development, Logos International Headquarters, 333 Elm Street, Dedham, MA 02026, U.S.A. Tel: (617) 326-7600; Fax: (617) 326-9341.

---

## **MITRE and MT Research**

The AI Center of MITRE's Washington D.C. site has advertised for two computational linguists to join their projects on Japanese-English MT and on the discourse component of a multi-modal user interface.

---

### **Dutch-English system**

In a recent issue of *Language International* (vol.4/3 p.32) it is reported that a company set up to provide specialist language services to the chemical industry (Hook and Hatton, Ltd. of Northampton, UK) has developed a simplified MT system for translating engineering specifications written in "very simple language". The program "has been specifically designed to translate from Dutch into English and 'to be aware' of the particular differences between the two languages." The developer Terence Lewis described the program as being "based on traditional approaches described in the literature, although some aspects, such as homograph resolution, have been built on known AI routines. The program uses a large dictionary of phrases and even complete sentences taken from a corpus of Dutch scientific and technical documents and a single word dictionary in which syntactic and some semantic information is provided.... At present the system's database enables it to deal with general scientific texts or texts in the fields of chemical engineering and computer science. The program runs on a 33 MHz 386 PC with 4 Mb or RAM and a 200 Mb hard disk."

---

## **USERS' VIEWS**

### **German MT Users' Group**

*Tom C. Gerhardt*

The aim of the group *Anwender Arbeitskreis Maschinelle Übersetzung* (Working group of machine translation system users) is to formulate the user's viewpoint with regard to machine translation, machine aided translation and other translation tools. The working group was initiated conscious of the lack of representatives and presentations of the users at the MT SUMMIT II in Munich 1989.

The first meeting at the 'Fraunhofer Institut für Systemtechnik und Innovationsforschung' in Karlsruhe (Germany) was a success: more than 30 persons, most of whom were users of MT systems, came and stressed the need for a system-independent working group with the aims:

- information exchange among the users
- building a platform for the articulation of users' interests
- representation of users interests with respect to third parties
- transfer of knowhow

In order to achieve these goals the working group established a number of subgroups:

- job description (including education and vocational training)
- comparison and evaluation
- introductory processes
- secondary tools in the translating environment

The User Working Group is not legally organized. It meets twice a year, with the next meeting

taking place on 28 October 1992 in Saarbrücken (Germany) just before the EAMT Workshop at the Saarbrücker Technology Fair.

For more information contact: Tom C. Gerhardt, ITS c/o BGP, 26 rue Marguerite de Brabant, L-1254 Luxembourg. Tel: +352-349414. Email: tom@crpcu.lu  
or: Dr. Hans Billing, GMD-PTF, Dolivostr. 15, W-6100 Darmstadt. Tel: +49-6151-875.732;  
Fax: +49-6151-875.740

### **Sixth meeting of the Group in Hürth**

More than 35 persons from five European countries met in Hürth near Cologne (Germany) in May (11th to 12th) to attend the spring plenary meeting of the Working Group of MT System Users. The meeting, which was hosted by the Bundessprachenamt (Federal Department for Languages), focused in its scientific papers on the services of the Bundessprachenamt and the evaluation of MT systems in European industries and administrations.

Members of the Bundessprachenamt (Ferdi Schneider, Theo Back, Peter-Tronje Hagen) presented the services of the administration in an overview and then concentrated on the translation information system (ÜSIS), which is used to supervise the different translation jobs and to avoid duplicated work, and on the lexical information system (LEXIS) handling a big lexical database with more than 1.3 million entries. Kirsten Falkedal from ISSCO in Geneva (Switzerland) spoke about current projects of the Institute, concentrating on the evaluation of MT systems. Sylvie Regnier (Aérospatiale, Paris) presented two methods of evaluation for Systran, one which is based on a comparative analysis and the other related to corpora. Julian Cocks from Siemens Nixdorf in Paderborn (Germany) described a translation procedure incorporating a MT system. The clue of the approach lies in the intelligent preparation of the text and a partially automated postediting of the document with search-and-replace commands. Winfried Lenders (University Bonn) gave a status report on the Text Encoding Initiative (TEI), highlighting its major outcomes and advantages for the NLP community. Peter Clark, project manager of the COSINE Sub Project P3 at Logica in London gave an introduction into the awareness actions of his project answering questions on e-mail facilities, which do not belong to the standard communication facilities in European enterprises.

The work programme of the meeting consisted of meetings of the sub groups on: job description (including education and vocational training), comparison and evaluation, introductory processes, secondary tools in the translating environment

### **COSINE support for the Group**

The Working Group of MT System Users was selected out of 400 special interest groups in Europe to benefit from the COSINE project 'Support of International SIGs'. The project aims to promote the frequent use of international networks as fast, cheap and reliable communication and information medium.

The support concept is foreseen as providing help with any problems that may occur while adapting computer equipment to the needs of networks and while subscribing to and making use of local, national and international networks. It provides access to international information and directory services and promotes the use of IXI networks.

---

### **ATA commissions study on MT for PCs**

*Muriel Vasconcellos*

[Adapted, with permission, from the ATA Chronicle]

To judge from the flood of inquiries that pour into the Committee on Machine Translation of the

American Translators Association (ATA) – sometimes as many as three a day – MT for personal computers is a hot topic. Now that several translation packages are available on PCs and finally within the reach of independent translators, these professionals want to know what the technology can do for them, the kinds of software and the language combinations that are available, and the criteria that should be used in assessing the relative usefulness of the many products that are coming to the market.

This is an area in which solid information has been lacking. In view of the urgent need for orientation and perspective, ATA has commissioned a study on the subject, which will be carried out by Christine Miller, a consultant who heads a software consulting firm in the Washington, DC, area and brings to the task an unusual combination of experience on both mainframe and PC-based MT systems. The study will: survey the current MT offerings for PCs; provide a checklist of criteria that translators can use for drawing comparisons between these products and determining which ones might be of interest for their particular applications; and go through the paces of a typical benchmark exercise. The study will be limited to batch systems and will not include any terminology management software that does not automatically generate full-text translation.

As indicated elsewhere in this issue (see "MT to be Featured in San Diego"), the results of the study will be presented at ATA's upcoming 33rd Annual Conference in San Diego.

The study is already contemplating a good half-dozen PC packages. MT News International readers are encouraged to submit the names of lesser-known PC products they may be familiar with, or any advertising they have seen, together with the vendor's address if possible, and also to share with Ms. Miller any experiences they have had. Material should be addressed to her at 2020 Pennsylvania Avenue, N.W., # 234, Washington, DC 20006, USA.

---

### **An Invitation to MT Users**

In this issue, there are reports of the MT Users' Group in Germany and announcements of conferences and seminars in which potential users can see MT systems in operation and present users can air their opinions. However, in *MT News International* the views of MT users have not been heard very much yet. This newsletter is intended as a forum in which they can recount their experiences, say what is wrong (or right) with the presently available systems, tell manufacturers and developers what their real needs are, and voice their hopes and expectations for systems which they would like to see on the market in the short and long term. If the researchers and developers are failing to examine fruitful lines, if they are ignoring or neglecting aspects which users think are important, then we would like to hear about them. Contributions from users of all kinds are welcome; they can be any length up to a maximum of 3000 words and in any form - letter, complaint, review - and can be sent to any of the editors listed on the front page. Let us hear your views!

---

## **PUBLICATIONS**

### **Directory of MT Systems**

In the last issue of *MT News International*, we announced the intention of IAMT to compile a directory of MT systems. During the next months, we shall begin the compilation of entries 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 directory will be ready (in preliminary form) for the fourth MT Summit next July. 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 the directory an accurate and reliable source for information about MT, and thus to promote interest in MT worldwide.

---

## **IEEE Expert**

Call for papers: *Special Track on Processing Natural Language*

**IEEE Expert** announces a Special Track on Processing Natural Language, edited by Terry Patten and Paul Jacobs. A Special Track is a collection of papers, united by a theme, running over several issues. For the Special Track on Processing Natural Language, papers are sought on the practical application of all areas of natural-language processing, including the following: computational morphology, parsing, semantic interpretation, pragmatic/contextual interpretation, discourse processing, text planning and realization/generation, and machine translation. An ideal paper would describe a technique or approach to natural language processing, along with an evaluation of how it works and how well it works in a particular application. Papers that give a critical evaluation are preferred over descriptions of programs or applications.

**IEEE Expert** is a magazine of applied AI, not a transactions or a journal. The magazine is a bridge between the research community and the user community. It aims to publish original papers that transfer to the user community ideas and tools that come out of research and development. Clear, not overly formal, writing is essential. Its readers are users, developers, managers, researchers, and purchasers who are interested in databases, expert systems, and artificial intelligence, with particular emphasis on applications. They want to learn about the tools, techniques, concepts, aids, and systems that have potential for real-world applications. Conceptual or theoretical papers are welcome, provided they clearly demonstrate their relevance to real-world applications – preferably through prototype implementations.

Submissions should be written according to the IEEE Expert style. The final articles should be about 8-9 printed pages, with about 10-12 references. All papers submitted will be carefully reviewed. Papers accepted on technical grounds are subject to copyediting by the Managing Editor's staff for clarity and expressiveness. Authors are asked to submit six copies (hard-copy only) of their paper by October 1, 1992 to: Terry Patten, Computer and Information Science, Ohio State University, 2036 Neil Ave. Mall, Columbus, OH 43210, U.S.A. Email: [patten@cis.ohio-state.edu](mailto:patten@cis.ohio-state.edu)

---

## **TRANSST: An International Newsletter of Translation Studies**

TRANSST, an international newsletter of translation studies (ISSN 0792-058X), is published by the M. Bernstein Chair of Translation Theory and the Porter Institute for Poetics and Semiotics, Tel Aviv University (Israel). It is edited by Gideon Toury, with the help of José Lambert (University of Leuven, Belgium).

TRANSST serves as an information clearinghouse for the Committee for Translation Studies of the International Comparative Literature Association (ICLA/AILC) and for the Scientific Commission on Translation and Interpreting of the International Association of Applied Linguistics (IAAL/AILA).

Editorial and administrative address: TRANSST, The M. Bernstein Chair of Translation Theory, Tel Aviv University, P.O. Box 39085, Tel Aviv, Israel. Bitnet: [TOURY@TAUNIVM](mailto:TOURY@TAUNIVM); tel. +972-3-6409502; fax: +972-3-6408980

---

## Natural Language Software Registry

The Natural Language Software Registry is a catalogue of software implementing core natural language processing techniques, whether available on a commercial or noncommercial basis. The current version includes speech signal processors, morphological analyzers, parsers, knowledge representation systems, multi-component systems, applications programs.

This document is available on-line via anonymous ftp to ftp.dfki.uni-sb.de (or: tira.uchicago.edu, IP 128.135.96.31), by email to registry@dfki.uni-sb.de, and by physical mail to the address below. If you have developed a piece of software for natural language processing that other researchers might find useful, you can include it by returning the description form, available from the same sources. Contact address: Elizabeth Hinkelmann (hinkelma@dfki.uni-sb.de), NL Software Registry, Deutsches Forschungszentrum für Künstliche Intelligenz, Stuhlsatzenhausweg 3, D-W-6600 Saarbrücken, Germany. Email: registry@dfki.uni-sb.de

---

### SCHOLAR Online Service for Text Analysis and NL Applications

SCHOLAR is an online information service covering all aspects of natural language processing in such fields as literary studies, linguistics, history, machine translation and lexicography. It consists of information such as book reviews, project reports, database listings, a conference calendar, and news of hardware and software relevant to the field. SCHOLAR is distributed occasionally as the quantity of information received allows. Contributions should be sent to Joseph Raben <jqrqc@cunyvm.cuny.edu>.

SCHOLAR is sponsored by Queens College and the Computer Center of the City University of New York, and is funded by the Andrew W. Mellon Foundation. To subscribe, send a message to listserv@cunyvm.cuny.edu, consisting of the following line only: subscribe scholar <your name>. To retrieve SCHOLAR's entire first release send mail to <listserv@cunyvm.cuny.edu> with the following as the body of the message: Get AZ Package

We note with interest that the first release contains, among other items, abstracts from Machine Translation, vol. 6 no. 1 (135 lines), vol. 6 no. 2 (151 lines), and vol. 6 no. 3 (229 lines), as well as a calendar of forthcoming events related to natural language processing (249 lines).

To retrieve an index of SCHOLAR files via anonymous ftp, use the following procedure:  
ftp jhuvvm.hcf.jhu.edu (or: ftp 128.220.2.2, if your computer does not recognize the full name of the remote host)

At the login prompt, type the word 'scholar'. At the password prompt, type your login user identification (this can be your full name, or the name of the account from which you have invoked the ftp utility). Once logged in, enter the following command to retrieve the SCHOLAR index file:

```
get index.scholar
```

If you wish to retrieve the entire release, enter the following command: mget AZ\*.\* When you have finished, type 'quit' to exit.

---

## Recent Publications

### *Books*

*Computing in linguistics & phonetics: introductory readings*, edited by Peter Roach. London: Academic Press, 1992. viii, 115 pp. ISBN: 0 12 589340 X. £14.95  
Includes (p.85-96) a chapter on Machine Translation by Geoffrey Sampson.

*Computers in translation: a practical appraisal.* Edited by John Newton. London, New York: Routledge, 1992. xx,238 pp. ISBN: 0 415 05432 X. Price: £37.50

Contents: Introduction and overview (John Newton). - The story so far: an evaluation of machine translation in the world today (Jeannette Pugh). - Made to measure solutions (Annette Grimaila and John Chandiooux). - The Perkins experience (John Newton). - Machine translation in a high-volume translation environment (Muriel Vasconcellos and Dale A.Bostad). - Esperanto as an intermediate language for machine translation (Klaus Schubert). - Limitations of computers as translation tools (Alex Gross). - Computerized term banks and translation (Patricia Thomas). - The translator workstation (Alan Melby). - SYSTRAN: it obviously works but how much can it be improved? (Yorick Wilks). - Current research in machine translation (Harold L.Somers). - Bibliography. - Glossary of terms. - Index.

The book is described by the editor as "a collection of practical accounts from researchers, system developers, system operators and translators, together with an appraisal of MT's position in the world today and an outline of the main avenues of current research" with due consideration "to 'lower-level' computerized translation tools ... which can offer tangible benefits in almost every area of written translation." The collection is thus a mixture of descriptions of some well-known systems in practical operation (Meteo, Weidner, PAHO, Systran), of MT history and current research activity, of term banks and translator workstations, and assessments of the limitations and possibilities of computer-based translation technology. Its coverage is not claimed to be exhaustive in any sense, but it is hoped to meet the needs of readers seeking information about MT in an accessible form. Given the paucity of such literature in this field, the volume is to be welcomed.

#### *Journal articles*

**Language Industry Monitor** , no.9 (May-June 1992), p.1-2: Systran revitalized? no.10 (July-August 1992), p.6: New MT system for Caterpillar now under development. - p.6: Japanese to demo Asian interlingua MT system this fall. - p.7: PC-based translators: they keep on coming [Sun Data Systems' Vertaler]

**Language International**, vol.4 no.3 (June 1992), p.32: New Dutch-English system [from Hook and Hatton Ltd., Northampton, UK; details elsewhere in this issue]; vol.4 no.4 (August 1992), pp.10-20, 29-35: Provo special (Geoffrey Kingscott) [including reports on the development of ALPNET, Weidner, ECS, TEI, WordPerfect, etc.]

**Machine Translation** vol.6 no.4 (1991/92), p.247-263: BITS: a Hebrew-English bibliographic translation system (Iris Arad). - p.265-287: Translator knowledge base for machine translation systems (Zenshiro Kaeasaki, Fumiyuki Yamano and Niri-yuki Yamasaki). - p.279-316: Resolution of ambiguity in Chinese and its application to machine translation (Chao-Huang Chang and Gilbert K.Krulle). - p.317-322: Book reviews (Christoph Zähler, J.C.Sager).

**Meta** vol.37 no.1, 1992. - pp.97-112: ETAP-2: the linguistics of a machine translation system (Jurij D. Apresjan, Igor M. Boguslavskij, Leonid L.Iomdin, Alexandre V.Lazurskij, Vladimir Z.Sanikov, Leonid L.Tsinman).

**British Computer Society, Natural Language Translation Specialist Group. Newsletter 20** (Spring 1992). Includes brief items on EAMT, machine translation in the UK, research at University of Essex, machine translation in Europe, machine translation in the USA, system

news (PARS/M and GEISHA, TECM, LOGOS), abstracts of talks ('Eurotra and other recent developments in machine translation' John Hutchins; 'Interactive machine translation for monolinguals' M.McGhee Wood; 'Machine translation using the SRI Core Language Engine' Manny Rayner). [Information about the group: J.D.Wigg, BCS-NLTG, 72 Brattle Wood, Sevenoaks, Kent TN13 1QU, UK.]

#### *Conference proceedings*

*Natural Language Processing Pacific Rim Symposium (NLPRS '91)*, November 25-26, 1991, Singapore. [Proceedings]. Sponsored by Information Processing Society of Japan, co-sponsored by Japan-Singapore AI Centre. [Singapore: Japan-Singapore AI Centre, 1991.]

Contents include: English to Korean transfer in MATES/EK (Heran Lee, et al.) - The future of machine translation: Dictionary Center and MT-VAN (Keiichi Nobuta) - Kanta: Malaysia's computer translation project (Ahmad Zaki Abu Bakar) - Machine translation technology and evaluation methodology (Hirosato Nomura) - Generation of polite expressions in multilingual translation (Susumu Akamine, et al.) - The generation of Chinese text in Japanese-Chinese machine translation system (June-Jei Kuo, et al.) - Korean generation in MATES/EK (Myung-Gil Jang, et al.) - Everything has not been said about interlinguae (Alain Polguère).

[Copies can be ordered from: Saw Ken Wye, Deputy Director, Japan-Singapore AI Centre, 75 Science Park Drive #01-01/04, CINTeCH II, Singapore 0511, Republic of Singapore. Fax: +65 779-6397]

*Third Conference on Applied Natural Language Processing*. Association for Computational Linguistics. Proceedings of the Conference, 31 March - 3 April 1992, Trento, Italy.

Contents include: Zero pronoun resolution in a Japanese to English machine translation system using verbal semantic attributes (H.Nakaiwa, S.Ikehara) - Compound nouns in a unification-based MT system (P.Bouillon et al.) - XTAG, a graphical workbench for developing tree-adjointing grammars (P.Paraoubek et al.) - Multi-purpose development and operation environments for natural language applications (Sergei Nirenburg et al.)

*30th Annual Meeting of the Association for Computational Linguistics*. Proceedings of the Conference, 28 June - 2 July 1992, University of Delaware.

Contents include: Examining upper and lower bounds on the performance of word-sense disambiguation programs (William Gale et al.) - A parameterized approach to integrating aspect with lexical semantics for machine translation (Bonnie J.Dorr) - Spatial lexicalization in the translation of prepositional phrases (Arturo Trujillo) - Metonymy: reassessment, survey of acceptability, and its treatment in a machine translation system (Shin-ichiro Kamei and Takahiro Wakao).

*Fourth International Conference on Theoretical and Methodological Issues in Machine Translation, TMI-92*. Proceedings of the Conference, June 25-27, 1992, Montréal, Canada, CWARC. [For details see report in this issue.]

[Copies may be ordered from: CCRIT, TMI-92, 1575 boul.Chomedey, Laval (Québec), Canada H7V 2X2. Email: tmi@ccrit.doc.ca]

*International Workshop on Fundamental Research for the Future Generation of Natural Language Processing (FGNLP)*. Proceedings, 30-31 July 1992, Hotel Dominion, Manchester. Sofia Ananiadou, editor. Manchester: Centre for Computational Linguistics UMIST. [For details see report in this issue.]



[Copies are available from: Kate Stacey, CCL, UMIST, PO Box 88, Sackville Street, Manchester M60 1QD, United Kingdom.]

*Translating and the Computer 13: The theory and practice of machine translation - a marriage of convenience?* Papers presented at a conference ...28-29 November 1991. London: Aslib, 1992. 165 pp. ISBN: 0 85142 295 0

Contents include: Why computers do not translate better (W. John Hutchins). - Evaluating MT systems (Jacques Durand). - From evaluation to specification (R. Lee Humphreys). - The Globalink Translation System GTS (Nigel Burnford). - SYSTRAN, or the reality of machine translation (François Secheresse). - Alpnet and TSS: the commercial realities of using a computer-aided translation system (Thomas Seal). - Current practical machine translation systems in Japan and future directions (Makoto Nagao). - EUROTRA: an assessment of the current state of the EC's MT programme (Doug Arnold and Louisa Sadler). - Machine translation seen as interactive multilingual text generation (Harold L. Somers)

---

*Items for inclusion in the 'Recent Publications' section should be sent to the Editor-in-Chief at the address given on the front page. We rely on readers to notify us of their publications.*

---