

MT News

International

Newsletter of the International Association for Machine Translation

ISSN 0965-5476

Issue no.7, January 1994

IN THIS ISSUE:

Conference Reports

Association News (AAMT)

People on the move...

Systems and Projects

Readers' Forum

Publications Announced and Received

Conference Announcements

Forthcoming Events

Advertisement

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@crpculu.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

Francophone Conference on MT held in Montreal 30 September, 1-2 October 1993

John Hutchins

The University of Montreal was host from 30th September until 2nd October 1993 to the **Troisièmes Journées Scientifiques** organised by the **Réseau Lexicologie, Terminologie, Traduction** of UREF (Université des Réseaux d'Expression Française). Previous 'journées scientifiques' have been held in Fez (Morocco), 20-22 February 1989, and in Mons (France), 25-27 April 1991. [For details of the proceedings of the Mons meeting, published as *L'Environnement traductionnel: la station de travail du traducteur de l'an 2001*, see 'Publications Received'.] The theme of this third conference in the series was *Traductique TA-TAO, recherches de pointe et applications immédiates*.

After welcoming speeches by the director of UREF, the rector of the University of Montreal, and by the conference organizer André Clas, the first session began with a scene-setting presentation by John Hutchins, who described the most recent developments in MT research. Virtually all facets of current MT activity were in fact represented during the conference. Statistical and corpus-based research was the focus of papers by Jean-Marc Langé (IBM France), Eva Dauphin (Aérospatiale), and Susan Warwick-Armstrong (ISSCO, Geneva). Current concerns with lexicon structure and acquisition were the topics of papers by Deryle Lonsdale (Carnegie Mellon University), Marie-Claude L'Homme (Lexi-tech, Canada), and Nina Leontyeva (Russian Academy of Sciences). Various problems of rule-based linguistic approaches were the subject of papers by Dominique Estival (ISSCO, Geneva) on unification grammar, by Laurence Danlos (Univ.Paris) on the treatment of constructions with *faire*, by Gaston Gross (Univ.Paris XIII) on conjunctions, and by Siety Meijer (Univ.Essex) on prepositions and their problems. Interlinguas and knowledge bases were dealt with by Gilles Serasset (Univ. Joseph Fourier) and Sergei Nirenburg (Carnegie Mellon Univ.) Computational issues were treated by Jean-Yves Morin (Univ.Montreal) on parsing formalisms, and by Mathieu Lafourcade (Univ. Joseph Fourier) on the LEAF neuronal network model.

In fitting with its Montreal location it was not surprising that four topics in particular were prominent in the conference: sublanguages, multilingual generation, evaluation, and the Meaning-Text model. Sublanguages were the theme of many papers, notably from Pierrette Bouillon (ISSCO, Geneva) on the IFENA system for avalanche bulletins, and (with Graham Russell) on general problems of sublanguages; and the quantitative typology of sublanguages was described by Sveta Sheremetyeva (Univ. Chelyabinsk, Russia). Multilingual generation from databases was treated by Liesbeth Degand (UCL Belgium) and Richard Kittredge (Univ.Montreal); and the evaluation of MT was discussed by Margaret King (ISSCO, Geneva) and Adriane Rinsche (Language Technology Centre, London). As for the 'meaning-text model', the conference had the honour of a stimulating presentation of lexical functions and MT by the originator of the approach himself, Igor Mel'chuk (Univ.Montreal), and a treatment of communicative structure in the MTM framework by his wife, Lidiya Iordanskaya (Univ.Montreal).

Presentations devoted to specific research programmes were provided by Alexander Lazourski (Russian Academy of Sciences) on the development of the Russian-English and English-Russian ETAP-3 system, by Eric Wehrli (Univ. Geneva) on his ITS-2 system for monolingual users, and by Mutsuko Tomokiyo (ATR Research Laboratories) on the ASURA speech interpretation system under development at ATR. Terminological questions were the subjects of a substantial number of papers, by Alan Melby (Brigham Young Univ.), Pierre Lerat (Univ.Paris XIII), Marc Van Campenhoudt (ISTI Belgium), Ingrid Meyer (Univ.Ottawa), Elisabeth Blanchon (INaLF-CNRS France), Patrick Burkhard (Union Bank of Switzerland), and Hussein Habaili (Univ.Tunis).

The future direction of MT and computer-based translation research was dealt with in depth by Christian Boitet (Univ. Joseph Fourier), who argued specifically for the investigation of dialogue-based models. The particular Canadian context of MT and the language industries was covered by contributions by Alain Maret and by Jean Quirion (both of the Canadian Secretariat of State.) And the concluding address from Jean-Claude Lejosne (Univ. Metz) put the development of translation and language industries into a broader social and cultural framework.

The conference was complemented by an exhibition of products from the Centre d'Innovation en Technologies de l'Information (TransSearch), CLC Limitée (Term Cruncher), Logiciels Tradulog (Proterm), Canadian government services (the PTT workstation for translators, and Termium on CD-ROM), Office de Langue Française (BTQ, the Quebec terminology databank), Linguistic Products (PC-Translator), Documensa (EdiBase), Bytetown Consulting (Mercury/Termex), IBM Deutschland (TranslationManager/2), Traductix (ATAO pretranslation system), John Chandioix Experts-Conseils (Météo, Le Général TAO, and TA-AU-CN), Socatra (XLT system), and the Observatoire Québécois des Industries de la Langue.

During the course of the conference, UREF and the Montreal University Press launched their new publication *La Traductique*, edited by Pierrette Bouillon and André Clas [for details see the section 'Publications Received' in this issue.] This important French collection of articles on all

aspects of MT and translation aids is a further sign of the vigour and enthusiasm within the Francophone world, which was indubitably demonstrated in this successful and informative conference in French-speaking Canada.

Workshop on Very Large Corpora 22 June 1994

Ken Church

The first annual workshop on "Very Large Corpora" was held on June 22, 1993, just before ACL-93, at the Ohio State University, Columbus, Ohio, USA. There was a special emphasis on "Academic and Industrial Perspectives" in order to take advantage of the location of the meeting, conveniently near three suppliers of information retrieval products: Chemical Abstracts, Mead Data Central (MDC) and Online Computer Library Center (OCLC). The workshop was sponsored by these three companies and by the ACL.

There was some good news and some bad news. The good news is that text is more available than ever before, and consequently, it is easier to use corpus data more effectively than it was in the 1950s, the last time that empiricism was in fashion. Many research laboratories have several million words of text, and a few places such as IBM even have a billion words, as we learned from Robert Mercer in his invited talk. All of this data provides a great opportunity, as evidenced by all of the recent activity in Europe, Asia and America. A third of the papers were from Europe, a third from Asia and a third from America.

On the other hand, the bad news was that text is more available than ever before. I used to think that a billion words was a lot of text, but after hearing Mark Wasson's invited talk, I'm not so sure anymore. Mark Wasson described the size of Mead's holdings, which are literally astronomical. Even the astronomer Carl Sagan would be impressed with Mead's billions and billions of words of text. The problem with so much text is that we can't do very much with it. Obviously, we can't afford to pay staff to read it all. Nor can we afford to process it with traditional slow/fragile NLP methods. Even text analysis methods such as so-called robust parsing may be lacking in industrial strength.

The workshop was a useful first step. The discussion pushed the researchers to think about even larger corpora, and the industrialists to think about somewhat more ambitious analysis techniques.

Next year's workshop will be held in Kyoto just before Coling-94. There will be a special emphasis on parallel text, a theme that may be of interest to many of the Coling-94 participants, especially those working on machine translation. [For further details see the section on Conference Announcements in this issue.]

Summer School on Computational Linguistics in Bulgaria 4-10 September 1993

Ruslan Mitkov

The Annual International Summer School "Contemporary Topics in Computational Linguistics" is an educational and scientific event of high international reputation. Lecturers of the Summer School are outstanding scholars from Germany and other Western countries. This year the summer school lecturers were leading academics from Germany, Canada, France, United Kingdom, Holland, Switzerland, and Spain. The survey courses given on the "hottest" topics of Computational Linguistics and Artificial Intelligence have made the summer school world known. From this year on the European Association for Machine Translation and the ECCAI have taken over the honorary sponsorship of the summer school. The summer school was organized for a third consecutive year in Bulgaria and welcomed participants mostly from Eastern Europe - above all from Bulgaria, Rumania, and Macedonia.

The summer school was the host of 8 lecturers and 30 participants (the participants were from Bulgaria, Belgium, France, Greece, Macedonia, Rumania, Russia, Switzerland, Spain; some of them were able to attend it thanks to the grants we managed to provide for them). The aim of the school is to familiarize young scientists from Bulgaria and Eastern Europe with recent scientific developments in the field of computational linguistics and artificial intelligence through state-of-the-art presentations by leading experts and scientists from Western countries.

This year's summer school took place from 4 September to 10 September at the KZU guest house, Vassil Kolarov Lake, Rhodope Mountains, Bulgaria. The scientific program included the following lectures: Formal systems, grammars and languages (Manfred Kudlek, Hamburg University, Germany); Sublanguage-based machine translation (Jun-ichi Tsuji, CCL, UMIST, Manchester); Multi-lingual text generation using Meaning Text language models (Richard Kittredge, Université de Montréal); Transfer or interlingua? (Dominique Estival, ISSCO, Université de Genève); The role of grammar formalisms in machine translation (Zaharin Yusoff, University of Science Malaysia, Penang); Text generation (Michael Zock, LIMSI, CNRS, France); New trends in mathematical linguistics (Carlos Martan-Vide, Universidad de Barcelona/Tarragona); Anaphora resolution (Ruslan Mitkov (University of Science Malaysia, Penang; Institute of Mathematics, Sofia).

Due to the great interest displayed by the participants in the school, the following additional courses were given: Meeting today's translation needs (Zaharin Yusoff); Mathematical model for personal pronouns, Mathematical models for time and aspect (Manfred Kudlek). Moreover, Michael Zock and Ruslan Mitkov were asked to give a couple of additional lectures for the participants of the Bulgarian National Seminar on extracurricular activities on MT (high school students and teachers), which took place in connection with the Summer School.

The scientific director of the Summer School is Dr. Ruslan Mitkov (Institute of Mathematics, Sofia) and Mr. N.Nikolov (Incoma's president) is the administrative director of the Summer School. The Summer School was co-organized by Dr. Michael Zock.

SwAN21 in Geneva 6-9 September 1993

Sandra Manzi

The SwAN21 Symposium was held in Geneva last September 6th-9th, in celebration of the 21st anniversary of the foundation of ISSCO, the Dalle Molle Institute for Semantic and Cognitive Studies.

A number of past researchers of the Institute came to share their experience and ideas in this occasion: we heard from them and from the present ISSCO personnel of progress made in AI and in getting AI closer to real needs, the re-thinking of certain techniques in Machine Translation, and the range of possible tools to be developed as translation aids, or for dictionary consultation.

Natural Language Processing was, as might be expected, the focus of most talks. Variety came from the multiplicity of facets within that general field, from pragmatics to text retrieval, from evaluation of NLP products to description of linguistic environments, the range was large. The atmosphere was very much that of a workshop, with participants, whether or not they made presentations themselves, challenging the speakers with questions and debate, as well as suggesting useful comments.

Speakers were (in timetable order): Manfred Wettler (Paderborn University, Germany): Word-associations and contiguities of words in texts. -- Dominique Estival (ISSCO): Grammars and reusability. -- Graham Russell (ISSCO): Environments for linguistic programming. -- Roger Schank (Northwestern University, USA): Why I don't do natural language anymore. -- Jan Hajic (Charles University, Czech Republic): Multilingual morphology with broad coverage. -- Dominique Petitpierre (ISSCO): Dictionary encoding for on-line consultation. -- Susan Armstrong (ISSCO): Multilingual corpora, a new resource for NLP and MT. -- Martin Kay (Xerox Parc, USA): On rehabilitating interlingual machine translation. -- Anne de Roeck (Essex University, UK): Towards a

formal model of pragmatics. -- Yorick Wilks (Sheffield University, UK): Whatever became of preference semantics? --Michael Hess (Koblenz University, Germany): Content-based text retrieval: an experiment in variable-depth NLP. -- Afzal Ballim (ISSCO): Modelling attitudes in dialogue. -- Pierrette Bouillon (ISSCO): Towards translation of avalanche bulletins. -- Phil Hayes (Carnegie Group, USA): Getting NLP out of the lab. -- Eric Wehrli (University of Geneva, Switzerland): Interactive translation: problems and solutions. -- Rod Johnson and C.J. Rupp (IDSIA, Switzerland): Constraints in language processing. --Margaret King (ISSCO): Whys and hows of evaluation

There should be proceedings available in the next few months: please contact ISSCO if you are interested. (54, route des Acacias, CH-1227 Geneva, Switzerland; email <issco-admin@divsun.unige.ch>).

AMTA Takes a Close Look at "Text and MT" 5 October 1993

Muriel Vasconcellos

AMTA's workshop "Text and MT," held in Philadelphia on 5 October 1993, drew an audience of lively and attentive participants looking for ways to make the best use of MT and predict the translation quality of particular MT systems. The workshop, co-organized by Claudia Gdaniec and Muriel Vasconcellos, focused on two main aspects of this challenge: characterizing the translatability of existing texts, and using tools that either critique texts or help authors to prepare copy so that it is fully MT-compatible from the outset.

In a general introduction to the characteristics of texts, Vasconcellos offered a series of criteria that determine the complexity of a text and help to explain why some are inherently more "translatable" than others. Veronica Lawson, a London-based consultant, described the characteristics of particular sublanguages with special emphasis on the sublanguage of patents. Gdaniec, of Logos Corporation, proposed a translatability index (TI) that would allow an MT user to determine whether a given text is suitable for machine processing and guide the user in the optimal preparation of such text for MT.

Sublanguages, of course, are more tractable fodder for MT systems. A methodology for quantifying the characteristics of an MT sublanguage was presented by Svetlana O.Sheremetyeva of Chelyabinsk State University in Russia, who showed how she used computer-aided analysis to arrive at a limited set of grammatical structures and verb forms that preferentially occur in a given sublanguage.

In a practical vein, Ghassan Haddad, senior manager of product translation at Intergraph Corporation, offered a definition of what constitutes good MT input based on his experience as a user of the company's proprietary MT system and went on to discuss methods and tools for aiding developers, current users, and potential customers in testing input quality.

Three of the speakers gave live demonstrations. Arthur Lee of Groupe Bull in France gave an interactive presentation on the use of Max, a product of Smart Communications, for controlling English input for MT in a large-scale production environment. Also on the subject of controlled English for technical documentation, an on-line presentation by the Carnegie Group's Phil Hayes and Steve Maxwell featured an authoring system based on the constraints of Caterpillar Technical English. This product is to be bundled with the MT system being developed for Caterpillar, Inc., by the Carnegie Mellon University's Center for Machine Translation. The workshop concluded with Sergei Nirenburg's presentation of CMU TWS, the interactive translator's workstation that has also been developed at Carnegie Mellon University.

Seminar on Matching Up MT Users and MT Environments 6 October 1993

Karen Spalink

The pre-ATA conference seminar "Matching Up MT Users and MT Environments" at the Adams Mark Hotel in Philadelphia on October 6 offered a new perspective on the MT selection process. Instead of surveying available MT systems in order to determine their usefulness for the documents at hand, the seminar looked at documents at hand in order to determine what parameters would make them suitable or unsuitable for MT. The seminar's approach was also different from previous seminars in that it addressed different types of users, namely free-lance translators, translation agencies, and in-house translation departments rather than users, vendors, and developers.

The morning session consisted of three presentations. Tony Roder, physicist and free-lance translator from California, gave the lead-in speech. He encouraged the audience to accept the changes that are in store for the translation profession. Gregory Zaretzky, president of the corporate word, inc., expressed his reservations about the usefulness of currently available commercial MT systems in a translation agency environment. Ghassan Haddad, senior manager of the documentation department at Intergraph Corporation, explained how his company, in effort to "write with MT in mind", successfully produced controlled English input while preserving a natural language flow in the original source text. Veronica Lawson, independent MT consultant based in London, put the various aspects into the broader MT perspective.

A new format was used in the afternoon session. Participants were assigned to small groups. Each group was presented with a set of original source documents. The group members had to identify text and non-text parameters that would determine whether text was suitable for machine translation. Besides producing concrete results that can be used in the decision process on the selection of an MT system, the group work presented a great opportunity to network and share views, opinions, and experiences.

The analysis of the evaluation forms indicates that most participants thought the seminar to be informative and productive and that more events like this should be planned in the future.

Translating and the Computer 15 **18-19 November 1993**

John Hutchins

The year's Aslib conference was distinguished by four presentations specifically devoted to MT. As in previous years the conference was held at the CBI Centre in central London and provided translators and others interested in the language industries with a wide range of presentations covering most aspects of the use of computers in translation work. Speakers covered problems of terminology control (Patrick de Gale, Sage de Gale Information Systems), terminology processing and management (Blaise Nkwenti-Azeh, UMIST), developing a management package for translators (Peter Barber), computer-based terminology applications in the Commission of the European Communities (Roger Bennett), collaborative authoring and hypermedia documents (Jörg Haake, Integrated Publication and Information Systems Institute, and Christine Neuwirth, Carnegie Mellon University), resources on Internet of interest to translators (Gerhard Obenaus, University of Iowa), dealing with foreign languages using WordPerfect (Peter Kahrel), and bilingual lexicography at Oxford University Press (Valerie Grundy and Marie-Hélène Corréard).

More directly relevant to MT as such were the presentations by Arthur Lee on controlled English whether for MT or not -describing the implementation at Bull SA (France) of the Max system from Smart Communications [see also separate item in this issue] - by Donia Scott on research at Brighton University for the development of computer tools for drafting multilingual documents, by Richard Birch (Rank Xerox) on what translators will be looking for in translation workbenches in the future, and by Norman Fraser (Vocalis Ltd.) on the SUNDIAL project (funded by ESPRIT) on the design of spoken language interfaces for telephone access to computer databases by members of the general public.

Two of the four presentations specifically devoted to MT inaugurated the conference on the first day. Yorick Wilks (University of Sheffield) gave a lively exposition of the statistical approach to MT being researched at IBM, highlighting its challenge to traditional rule-based MT, its

achievements and its limitations, and discussed the impact of current corpus-based research activity in the field. He was followed by Seamus Derrington (Nissan European Technology Centre), who in a philosophically oriented paper attributed the failure of MT to achieve high quality translations (FAHQ) to the lack of research on the foundations of translation and communication, an emphasis on short-term goals and the failure to fully integrate the various disciplines involved in language, cognition and computing.

On the second day, delegates heard the two other MT-specific presentations. Margaret King (ISSCO, Geneva) outlined the increasing activity in the area of evaluating MT systems and natural language processing software in general, explained the difficulties of reaching agreed methodologies and measurements and described the initial deliberations of the recently founded Expert Advisory Groups on Language Engineering Standards (EAGLES). She was followed by Veronica Lawson, who in a paper jointly written with Muriel Vasconcellos, gave the results of the latter's survey this year of MT users - perhaps the most wide-ranging ever, and certainly the most up to date [see MTNI#6, p.12-17]. Within the last two to three years, a substantial number of major users have been added to the companies already familiar in the field. In many cases the total number of words dealt with is in excess of a million a year - in one case, no less than 45 million! It is now estimated that some 300 million words are translated every year using MT systems. In addition, there is an indeterminable amount translated by non-translators with cheaper software for personal computers. Interesting comments by some of the "satisfied customers" of these cheaper systems served to illustrate that MT is beginning to be a genuinely mass product - a development, which perhaps many researchers and manufacturers have not yet faced up to seriously.

The papers of the conference [for details see 'Publications Received'] are available from Aslib, The Association for Information Management, 20-24 Old Street, London EC1V 9AP (Tel: +44 (0)71 253 4488)

Japan-US Workshop on Machine Aided Translation, 22-24 November 22-24

Eduard Hovy (ISI)

In continuing efforts to improve communication and scientific exchange between the Japan and US, a small three-day workshop on Machine-Aided Translation was organized by the Japan International Science and Technology Exchange Center (JISTEC) and the Science and Technology Agency (STA). The workshop, held at the US Department of Commerce building in Washington, DC, from November 22 to November 24, was attended by 10 senior MT researchers from Japan, 2 from England, and 18 from the US, as well as several observers from various Japanese and US Government organizations. The technical organizers were Makoto Nagao from Kyoto University and Jaime Carbonell from Carnegie Mellon University.

Takeo Arimoto, Director of the STA, opened the workshop, stressing the ever-increasing importance of MT and speech-to-speech translation. Patricia Yoshida gave the welcome address, describing US Government efforts in tracking MT. The program then alternated talks by US and Japanese participants.

In a paper by Teruko Mitamura and Eric Nyberg from Carnegie Mellon University (CMU), Teruko Mitamura gave an overview of the Interlingua representations used in KANT. Hozumi Tanaka from the Tokyo Institute of Technology described the incorporation of Japanese word segmentation into parsing by linking morphological and syntactic information and rules into an LR table, using EDR dictionary entries for this task. In a paper that received much later commentary, Satoru Ikehara from the NTT Laboratories described the extensive feature sets used to define lexicon entries for the NTT lexicon. Jaime Carbonell from CMU discussed aspects of MT evaluation, illustrating with a partial evaluation of KANT. Terumasa Ehara described the speech transcription, translation, and abbreviation process used by the Japanese Broadcasting Corporation NHK. Next, responding to a recent mischaracterization of one of her statements as being against sublanguages, Muriel Vasconcellos presented a plea for MT workers to define their terms. David Farwell of the

New Mexico State University described aspects of dialogue that bedevil traditional MT approaches. Aravind Joshi and Martha Palmer from the University of Pennsylvania described the use of lexicalized Tree Adjoining Grammar to perform Chinese to English translation of difficult verbs like "break". Makoto Nagao from the University of Kyoto described several heuristics useful for analysis, including the treatment of conjunctions, attachment problems, and determiners. He also announced the free availability of a new Japanese parser, the KN-parser (Kurohashi-Nagao parser) from the University of Kyoto (contact Sadao Kurohashi at kuro@kuee.kyoto-u.ac.jp). Sergei Nirenburg from Carnegie Mellon University described the current structure of the joint CMU-New Mexico-ISI system Pangloss and outlined the role within it of the new subproject Mikrococosmos. Eduard Hovy from the Information Sciences Institute of the University of Southern California followed this with a description of the current work in Pangloss to build an Interlingua Ontology incrementally. The day concluded with an elegant reception hosted by the STA.

On the second day, Kevin Knight of the Information Sciences Institute described statistically based work to insert articles into English text, useful not only for Slavic and Japanese to English translation (since these languages have no articles) but also for correcting non-native speakers' text or OCR input. Using material from her recent book, Bonnie Dorr from the University of Maryland described the use of LCS structures for translation and then outlined plans for the large-scale acquisition of verb information. Yuji Matsumoto from the Nara Institute of Science and Technology described the automated alignment of parses from Japanese and English and the subsequent extraction of transfer rules. Toshio Yokoi from the EDR project outlined the present and future status of the EDR dictionaries, which are now moving into the beta test stage. Rebecca Finch from the LDC at the University of Pennsylvania listed the LDC resources appropriate and available for MT research. Tatsuo Ashizaki from JICST presented the frequencies of use and their changes since 1990 of various types of entities in the JICST dictionaries. Next, Steven Helmreich from the New Mexico State University outlined work on the acquisition by automatic means of knowledge for both MT and TIPSTER (an information retrieval task). Philip Hayes from the Carnegie Group described the controlled language environment used in the system translating Caterpillar text. Hirosato Nomura from the Kyushu Institute of Technology described the overall JEIDA approach to evaluation, focusing on the various types of measures appropriate for various types of interests in MT. Hitoshi Iida from the ATR Laboratories described example-based translation of dialogue, especially techniques for measuring the similarity of two examples text fragments, one in each language. Lori Levin from Carnegie Mellon University described work in the new project, joint with New Mexico State, on translating spontaneous discourse.

The final day ended on a high note. Akira Kurematsu of the University of Electro-Communications presented an overview of speech translation, among others by the ATR-CMU-Karlsruhe/Siemens collaboration. Yorick Wilks from the University of Sheffield outlined some of the problems that have been and will increasingly be encountered in cross-country MT collaborations, entertaining the idea of "language-based interlinguas". Finally, Jun-ichi Tsujii from the University of Manchester described the role of language-independent ontological knowledge in a transfer system.

After a break, Makoto Nagao led a general discussion about new efforts for collaboration between Japan and the US, including aspects of data exchange, research visits, and so on. He then closed what by many accounts had been an unexpectedly interesting and informative workshop.

Natural Language Processing Pacific Rim Symposium 6-7 December 1993

The second Natural Language Processing Pacific Rim Symposium (NLPRS '93) was sponsored by the Special Interest Group on Natural Language Processing (SIGNLP) of the Information Processing Society of Japan (IPSJ) and jointly sponsored by Fukuoka Institute of Technology. It was held at the Fukuoka Institute of Technology on 6th and 7th December 1993.

The following papers were presented: A collocation-based transfer model for Japanese-to-Korean machine translation (EunJa Kim and Jong-Hyeok Lee, Korea); A computational grammar for Hindi: simple sentences and beyond (G.V. Singh and D.K. Lobiyal, India); A hybrid method for the resolution of prepositional phrase attachment problems (WonSeog Kang, Jungyun Seo and GilChang Kim, Korea); A method for detecting Japanese homophone errors in compound nouns (Masahiro Oku, Japan); A sublanguage approach to automatic term recognition (Sofia Ananiadou, UK); Algorithmic information extraction from natural language document (Shuichi Misumi, Hideaki Nagai, Teigo Nakamura and Hirosato Nomura, Japan); An experimental system for learning "color" concepts based on mental-image directed semantics theory (Seio Oda, Yasushi Nishimura, Masato Shiraishi and Masao Yokota, Japan); An interactive method for semantic disambiguation in sentences by selecting examples (Masaya Yamaguchi, Nobuo Takiguchi, Yoshiyuki Kotani and Hirohiko Nisimura, Japan); A lexical ellipsis-resolution mechanism: A corpus-based approach for ellipsis resolution (Ikuo Kudo and Mutsuko Tomokiyo, Japan); Anaphora resolution by machine in the SD-form semantics model (Masahiro Wakiyama, Guifeng Shao, Koichi Nozaki and Eiji Kawaguchi, Japan); Asking as a basis for processing the unknown terms, in a natural language user interface to a database (Zouheir Trabelsi, Yoshiyuki Kotani, Nobuo Takiguchi and Hirohiko Nisimura, Japan); Automatically extracting simple auxiliary phrases from a corpus (Hiroyuki Shinnou and Hitoshi Isahara, Japan); Chained functions structure: meaning representation for natural language interface (Yasuharu Namba and Hiroshi Kinukawa, Japan); Computer-assisted writing system for specification documents described in natural language (Katumi Tanaka, Koichi Nomura, Kimihito Takeda and Hideki Hirakawa, Japan); Conceptual transfer in an interlingua method and example based MT (Hiroshi Yasuhara, Japan); Coordination structures in law Sentences and a controlled linguistic model (Hideaki Iwamoto, Kaoru Nagai, Hidetoshi Nagai, Teigo Nakamura and Hirosato Nomura, Japan); Corpus-based modality generation for Korean verbs in PIVOT E/J/K system (Dong-Un An, Gil-Chang Kim, Jong-Hyeok Lee and Kazunori Muraki, Korea); Corpus-based Chinese text recognition (Xia Ying and Chang Xingong, China); Determination of sentence structure for understanding of scientific and technical documents (M. Takeda, N. Kusumoto, J. Suda and F. Matsuo, Japan); Dialogue-based MT for monolingual authors and the LIDIA project (Christian Boitet and Herve Blanchon, France); Disambiguation with distinctive features extracted from an example-base (Naohiko Uramoto, Japan); Document structure extraction for interactive full-text retrieval (Kazuo Sumita, Kenji Ono and Seiji Miike, Japan); Effect of discourse constraints for ellipsis supplement in Japanese texts (Shinji Fujisawa, Shigeru Masuyama and Shozo Naito, Japan); Epsilon [e]: tool kit for knowledge acquisition based on a hierarchy of pseudo-Text (Jun-ichi Tsujii and Sofia Ananiadou, UK); Example-based stochastic analysis of Japanese derivative words (Natsuki Uchimar, Teigo Nakamura, Yoshiaki Miyamoto and Toru Hitaka, Japan); Feature based TAG in place of multi-component adjunction: computational implications (B.A. Hockey and B. Srinivas, USA); Fragments of a Japanese appending X-bar grammar (AXG) (Bengt Sigurd and Birgitta Lastow, Sweden); Generation of elliptical subject by using bilingual statistical information for Korean-to-English machine translation (Byeongrak Seo and Yung Taek Kim, Korea); HPSG as constraint satisfaction problem (Philippe Blache, France); Handling non-Roman scripts in English based programs (Ramesh K. Verma, India); Heuristics for generating prominence in dialogues (Shozo Naito and Akira Shimazu, Japan); How to describe the meaning of natural-language sentences and system knowledge by using SD-forms (Eiji Kawaguchi, Masahiro Wakiyama, Koichi Nozaki and Guifeng Shao, Japan); Identifying deep grammatical relations in Korean relative clauses using corpus information (Jaehyung Yang and Yung Taek Kim, Korea); Indexing compound words from Korean texts using mutual information (Pan Koo Kim and Yoo Kun Cho, Korea); JPSG versus dependency grammar (Hisami Konishi Springer and Stanley Starosta, USA); Knowledge acquisition from natural sentences (Shimon Ando, Shou Kusumoto, Takeshi Uchida, Hidehiro Moriya and Makoto Masutani, Japan); New methods for deciding types of erroneous characters wrongly substituted, deleted and inserted in Japanese bunsetsu and for correcting these errors (Tetsuo Araki, Satoru Ikehara and Nobuyuki Tsukahara, Japan); Object oriented design of semantically driven tag based Hindi sentence analyzer (Om Vikas, Said Younes and Vijay Aggarwal, India); On correction

of ill formed Hindi sentences (Chinmoyee Sanyal and R.M.K.Sinha, India); On restoration system for the elliptical words of Japanese concatenated nouns (Masato Shiraishi, Yasushi Nishimura, Seio Oda and Masao Yokota, Japan); Pattern-based translation: context-free transducer and its application to practical NLP (Hiroshi Maruyama, Japan); Penstation for idea processing (Kenji Satoh and Kazunori Muraki, Japan); Processing of homonyms using co-occurrence relation between a noun and a verb (Masahito Takahashi, Takehiro Onomatsu, Kenji Yoshimura and Kosho Shudo, Japan); Prosodic information control based on lexical discourse grammar (Shinichi Doi, Shinichiro Kamei, Kazunori Muraki, Yukio Mitome and Kazuhiko Iwata, Japan); Repairing self-repairs in Japanese (Yuji Sagawa, Noboru Ohnishi and Noboru Sugie, Japan); Rules of rhythm construction in TANKA and judgment method of rhythm (Kenzi Imamura, Japan); SEG-TAG: a Chinese word segmentation and part-of-speech tagging system (Chao-Huang Chang and Cheng-Der Chen, ROC); Semantics of Japanese sentence final particles -- about 'Yo' 'Ne' and 'Na' (Susumu Ono and Hiroshi Nakagawa, Japan); Sentence generation from partially constrained feature structures (Koichi Takeda, Japan); Simple method for natural dialogue generation by eliminating redundant components (Toshihiro Omori, Hideaki Nagai, Teigo Nakamura and Hirosato Nomura, Japan); Structural misalignment in machine translation (John D. Phillips, Japan); Stochastic model for focus and its application to dialogue generation (Tsutomu Matsubara, Hideaki Nagai, Teigo Nakamura and Hirosato Nomura, Japan); Syllable-based morpheme isolation and morphological alternation of Korean words (Seung-Shik Kang, Korea); Technical terms in scientific and Technology-related documents (Yasuhito Tanaka, Japan); The design of MCP multiprocessor system (Kuanquan Wang, Nengzhong Wang and Weiqun Zhang, China); The DiCo approach to computational lexicography and its applicability to Chinese (Ngai-lai Cheng and Alain Polguère, Singapore); Two trie structures for natural language dictionaries (Katsushi Morimoto and Jun-ichi Aoe, Japan); Viewing dependency parsing as a statistically-based tagging process (Zhou Ming, Huang Changning, Zhang Min, Bai Shuanhu and Wu Sheng, China); Word sense disambiguation using distance vectors of words (Yoshiki Niwa and Yoshihito Nitta, Japan).

For further information: Hirosato Nomura (NLPRS), Dept. of Artificial Intelligence, Kyushu Institute of Technology, Iizuka, 820, Japan. Tel: +81-948-29-7613; Fax: +81-948-29-7601; Email: nomura@ai.kyutech.ac.jp

ASSOCIATION NEWS

Association for Machine Translation in the Americas Meeting of the General Membership Holiday Inn, Philadelphia, PA, Tuesday, 5 October 1993

AMTA President Muriel Vasconcellos called the meeting to order at 5:30 p.m. She announced that Secretary W. Scott Bennett was absent and that she had asked Deanna Hammond to act as secretary for the purpose of taking minutes.

Agenda

The proposed agenda was approved by those present.

Minutes

The minutes of the November 1992 membership meeting were read and approved.

Report of the President

M. Vasconcellos reported on the state of AMTA.

1. **Membership.** As of mid-September AMTA had 349 members, 149 of whom had failed to renew in 1993 but had nevertheless been carried on the rolls. Nineteen were corporate or institutional members. The printing and mailing of newsletters and the Yellow Book to unpaid members had represented a heavy cost to the Association in 1993. In order to prevent this from happening again, the Executive Committee had recommended in the future sending only the first issue of the

newsletter in the calendar year to those who have not renewed their membership. Those joining in the course of the year would receive all the back issues for the particular calendar year.

2. **Activities.** AMTA events during 1993 included the workshop "Text and MT" and a seminar "Matching Up MT Users and MT Environments" on 6 October cosponsored with the American Translators Association.

3. **Bylaws.** Vasconcellos reported that a permanent version of the Bylaws was in preparation and would be circulated to the membership in advance of the next General Meeting. Until 1994 the existing Provisional Bylaws were in compliance with IRS requirements, but a permanent version, approved by the membership, would be needed after that time.

4. **IAMT.** M. Vasconcellos had been chosen President-elect. W.Scott Bennett had replaced her as Secretary. A permanent set of bylaws had been prepared and were adopted at the meeting in July.

TMI

Sergei Nirenburg, chair of the steering committee for the International Conferences on Theoretical and Methodological Issues, reported on TMI-93, held in Kyoto in July. He announced that the next TMI would be held in the summer of 1995 in Louvain, Belgium, shortly before MT Summit V in Luxembourg. He mentioned that there had been discussion about whether TMI should be held at the same time as the Summit or during the years in between and asked for opinions on this matter. It was pointed out that scheduling the two conferences back-to-back helps to promote attendance for both. On the other hand, it was important that TMI not be relegated to the status of a "sideshow." On the whole the group felt that the two conferences are complementary and that there are advantages to holding them at the same time.

Report of the Treasurer

In the absence of R. Merchant, Monica Kusakabe presented the Treasurer's report. As of the end of September, AMTA had had an income of US\$33,650.82 and expenses of US\$27,258.17, leaving a balance on hand of US\$6,392.65. The President noted that it was becoming necessary to spend money to do tasks that had previously been being done by volunteers. For example, inquiries are received from around the world and a person is now being paid on an hourly basis to help handle the responses. Also, large mailings will be handled commercially in the future. To meet these additional costs, it is essential to "build up the coffers" and plan spending more carefully. It may be necessary to raise dues; the \$40 Associate Member fee barely covers the cost of preparing and distributing the newsletter. There will be another membership drive in 1994.

"Text and MT" Workshop

M. Kusakabe reported on the workshop that had just ended. M. Vasconcellos asked the participants for comments on the program, and the group agreed that the focus had been good. A suggestion was made to hold a business-oriented conference.

AMTA/ATA Seminar

K. Spalink reported on the cosponsored AMTA/ATA seminar to be held on 6 October. There had been a lot of interest in the event: it had been announced that registration would be closed off at 50 participants and a number of people has to be turned away.

Information Committee

J. Murgida, Co-chair, reported that she had set up a system for handling the inquiries received. Boilerplate letters had been prepared. They do not recommend products and or undertake to offer information on products available.

M. Kusakabe, Co-chair, reported that she had set up a lending service to provide members with access to one-of-a-kind reports, bibliographies, etc. She sends out materials on two-week loan and the borrower pays the return postage.

Newsletter and Advertising

Advertising Coordinator Bill Fry noted that the last issue of MT News International had announced that advertisements would be accepted starting with the January 1994 issue. Citing the current tone of MTNI, he felt there was need for some kind of criteria regarding the kind of ads that should be accepted. He asked the members present for guidance and comments. It was generally agreed that the

tone of the ads was important. One member said that he likes MTNI the way it is now, as a journal without ads. Another stressed that the Association should not become cheerleaders for particular products and that there should be a mechanism for ensuring that "not just anything" is accepted. It was pointed out that MTNI already publishes unbiased reports based on company press releases and the reader can keep up to date by following that section. The consensus was that ads would need to be monitored for both content and appearance and that the Editorial Board should establish a policy. The members felt it was essential that MTNI maintain an impartial position with respect to individual commercial products.

Yellow Book

B. Fry noted that the book had been put together in less than two months. The price was set at US\$6.00, which does not cover the cost. He acknowledged the contribution of Carmen Lozano as editor.

Future Activities

B. Fry expressed interest in holding a trade show with invitations to the press and potential users. Corporate sponsorship could be sought. It was agreed that a trade show was a good idea, keeping several considerations in mind, however. First, AMTA's main purpose is educational - i.e., to exchange information - and it should avoid any risk of being considered a trade association. Second, it was pointed out that exhibitors might have a limited number of shows they would be able to attend.

The group also discussed the possibility of holding events in conjunction with the annual meetings of the Association for Computational Linguistics or the Society for Technical Communication. Others felt it was time for AMTA to do a program independently. M. Vasconcellos suggested holding a conference every two years to alternate with the international MT Summits. There was general consensus that AMTA should have an independent conference in 1994. It was agreed to plan something on a fairly modest scale in 1994 and have a bigger conference in 1996.

EAMT

Veronica Lawson gave a report on EAMT, including the events that had been held, the elections, and plans for MT Summit V in Luxembourg in 1995.

Summit IV

Claudia Gdaniec reported on MT Summit IV, held in Kobe in July 1993.

International Fund

It was noted that Margaret King wanted to establish an international fund for people who had difficulty paying their dues, for example in the countries of Eastern Europe.

Adjournment

There being no further business, the meeting was adjourned at 6:30 p.m.

PEOPLE ON THE MOVE...

Winfield Scott Bennett has left Siemens and METAL and has joined Logos Corporation as a computational linguist. Scott can be reached at Logos Corporation, 200 Valley Road, Suite 400, Mt. Arlington, NJ 07856; tel. 201-398-8710, fax. 201-398-6102.

Readers of MTNI will have heard with sadness that *Don Walker*, without whom the Association for Computational Linguistics would not be what it is today, passed away in December 1993.

SYSTEMS AND PROJECTS

News from Smart Communications, Inc.

[Press release]

Smart Communications, Inc., of New York, NY, USA, a pioneer in pre-editing, has introduced a new product, *MaxTrans*. A tool for technical writers and translators, MaxTrans combines the following functions in a single workbench:

- * The Smart Expert Editor, alias "Max": Max is a text-critiquing tool designed to help technical writers stay within the bounds of controlled English. It has a basic technical vocabulary of 2,500+ keywords plus capacity for unique product nomenclature.

- * The Smart Translators: Offered as a suite of tools within an X-Windows environment, MaxTrans users can translate controlled English into French, Spanish, Italian, German, and Greek. An Asian language version translates controlled English into (simplified) Chinese as used in the People's Republic of China.

The MaxTrans software is written for workstations that use X-Windows, Motif, or OpenLook. The product is integrated with FrameMaker 3/4, WordPerfect 5.1 for Unix, and Interleaf 6. Target workstations include Bull DPX/20, IBM RS/6000, SunSparc Solaris 2.3, DEC Alpha OSF/1, and the PowerPC.

This new Smart offering targets the "electronic book" market, which is growing rapidly as companies rely on CD-ROM delivery systems. Controlled English makes for faster and more accurate database navigation.

MaxTrans' components already have an impressive track record. One large user of Max, Rockwell International's Automotive Division, has successfully rewritten hundreds of pages of technical materials in clear, concise English. Other major pre-editors are Boeing and Airbus Industrie. With the advent of NAFTA, Rockwell International is embarking on a program to produce manuals using the Smart Translator for Spanish.

Another Smart client, Groupe Bull of France, has moved to controlled English input combined with MT via MaxTrans as a means of controlling the quality of its documentation. Bull has installed MaxTrans on a central DPX/20 Unix server (equivalent to the RS/6000) capable of serving 500 users in multiple locations. Writers at five sites in France now follow the same vocabulary, which is mapped to the various languages handled by MaxTrans. For Groupe Bull, a major feature of MaxTrans is the transparent ability to read/write Interleaf, FrameMaker, SGML, Microsoft Word (RTF), and WordPerfect files within the same workbench.

For more information, contact Smart Communications, Inc., 885 Third Avenue, 29th floor, New York, NY 10022, USA; telephone (212) 486-1894, fax (212) 826-9775.

CELEX database available on CD-Rom

[Press release]

The CD-ROM, which contains the CELEX lexical databases of English (version 2.5), Dutch (version 3.1) and German (version 2.0), is now available for research purposes from the Linguistic Data Consortium for \$150. For each language, the CD-ROM contains detailed information on the orthography (variations in spelling, hyphenation), the phonology (phonetic transcriptions, variations in pronunciation, syllable structure, primary stress), the morphology (derivational and compositional structure, inflectional paradigms), the syntax (word class, word-class specific subcategorisations, argument structures), and word frequency (summed word and lemma counts, based on recent and representative text corpora) of both wordforms and lemmas (English: 52446 lemmas, 160594 wordforms; German: 50708 lemmas, 359611 wordforms; Dutch: 124136 lemmas, 381292 wordforms). Postscript files describe the available lexical information in detail.

The original Celex databases can be consulted interactively either by using the SQL*PLUS query language within an ORACLE RDBMS environment, or by means of the specially designed user interface FLEX. The databases on this CD-ROM have not been tailored to fit any particular database management program. Instead, the information is presented in a series of plain ASCII files in a UNIX directory tree that can be queried with tools such as AWK or ICON. Unique identity numbers allow the linking of information from different files. As in the original databases, some

kinds of information have to be computed on-line. Wherever necessary, AWK functions have been provided to recover this information. README files specify the details of their use.

The CD-ROM is mastered using the ISO 9660 data format, with the Rock Ridge extensions, allowing it to be used in VMS, MS-DOS, Macintosh and UNIX environments.

Anyone who would like to purchase the CD-ROM should send a check or purchase order made payable to the "Trustees of the University of Pennsylvania" to: Judith Storniolo, Administrative Assistant LDC, Linguistic Data Consortium, 441 Williams Hall, University of Pennsylvania, Philadelphia, PA 19104-6305; Tel: +1 215 898-0464; Fax: +1 215 573-2175; Email: storniol@unagi.cis.upenn.edu

Further details concerning the lexical databases can be obtained by anonymous ftp from the LDC as follows: ftp.cis.upenn.edu; directory: pub/lcd; file: celex.info.tar.Z

Persons outside Europe who are interested in CELEX but unable to retrieve and print the introductory text themselves, may request a hard copy of the document from the LDC. Persons in Europe who want a hard copy of the document mailed to them, and anyone who still has technical questions after reading the document, should direct their inquiries to: Richard Piepenbrock, CELEX Project Manager, Max-Planck-Institut für Psycholinguistik, Wundtlaan 1, 6525 XD NIJMEGEN, The Netherlands. Tel: (+31) (0)80-615797; Fax: (+31) (0)80-521213; Email: celex@mpi.nl

Software from PARS Ltd.

[Extracts from Information letter]

PARS 2.04 is an English-Russian-English system for translating scientific and business texts. PARS runs on IBM-compatible personal computers under MS-DOS, versions 3.1 and higher. In July 1993, PARS was chosen among several MT systems for implementation at the Izvestia Concern, Moscow, Russia, to translate information-new data base of the VWD agency, Germany. We also began developing a German-Russian MT system for the same user.

PARS can do the following:

- translate full texts or their parts
- use the specialized PARS text editor, marking words missing in the dictionary and synonyms of polysemantic words with different colors; the editor also makes it possible for the end-user to find the non-found words in the output text; change places for adjacent words; view the source and target texts either autonomously or in a synchro mode; mark blocks, format the texts, as well as do a number of other operations; the PARS text editor is a 2-window editor.
- work with 'external' text editors which support standard ASCII files;
- make use of up to 5 dictionaries in each translation session;
- create, extend and update the dictionaries, both autonomously and from the screen.

PARS is basically a direct-translation system, based on the Georgetown translation philosophy, with a lot of transfer routines. In the PARS system, the dictionary entry has a comparatively simple structure.

PARS 2.04 is supplied with three kinds of English-Russian-English dictionaries: 'near full', 'basic', and 'mini'. The list of the dictionaries as of September 1993 is the following:

- general basic: 14,000 word entries
- software basic: 9,000 entries
- electric engineering basic: 8,000 entries
- business correspondence and contracts basic: 6,000 entries
- medicine (immunology, obstetrics, venerology basic): 18,000 entries
- powder metallurgy basic: 3,500 entries
- chemistry mini: 2,000 entries
- terms on trademarks, near full: 1,000 entries

The following terminological dictionaries will be ready in October: economics (near full), mathematics (basic), agriculture (mini), machine building (basic), law (mini), patent terms (near full), social science (mini).

RDT (Resident Dictionary Tools) is a flexible and resident software package to run on IBM-compatible PCs under MS DOS as a translator's workbench. The end-user can apply RDT from any text editor that works with standard ASCII files.

RDT supports inverted bilingual dictionaries, such as English-Russian-English ones. Up to four dictionaries can be accessed in each session. RDT is supplied with the same dictionaries as PARS, as well as with:

(a) German-Russian two-way general basic and finance near full dictionaries of 15,000 entries each.

(b) Spanish-Russian two-way basic general, business, technical dictionary of 32,000 entries.

(c) English-Ukrainian two-way basic dictionaries of 25,000 general and political terms and 1,000 museum terms.

Detailed tables of Russian and/or Ukrainian morphologies can be supplied as Help facilities with RDT, as well as a systematized description of Russian morphology.

The **RUMP** system is supplied with a two-way Russian-Ukrainian basic dictionary of 17,000 general and political terms. The system comprises the same built-in text editor as the PARS system, which makes working with the two systems an easier task.

RUMP has dozens of users all over Ukraine. In particular, 20 copies have been installed at the Ukrainian Parliament.

For further information: M.S.Blekhman, PARS Ltd., Director: Lingvistica '93 Co., 94a Prospekt Gagarina, apt.111, Kharkov 310140 Ukraine. Tel: (0572) 27-71-35.

SPRINT System Available on PC

[Information leaflet]

The All-Russian Centre for Translation offers a PC version of its English-Russian machine translation system. System characteristics:

* Translation speed: 40,000 symbols (24 pages of typed text) per 20 min.

* Translation quality: not less than 80% of translated text is clear to those skilled in the art without reference to the original text.

* Source dictionary volume: more than 30,000 stems. Target dictionary volume: more than 65,000 words and word-combinations.

* Subjects: current English; current scientific English; computer science and programming; economics

* Computer type: IBM PC/AT-286 of standard configuration

* Disk memory: 4 MB.

* Operating system: MS-DOS 3.0 and higher

Address: All-Russian Centre for Translation, 117218 Moscow, Krjijanovskiy ul., Building 1, 14; Tel: (095) 124 36 74, or (095) 124 34 56.

Spanish Amigo for Windows

[From a press release]

WestCliff Software Inc. announce the appearance of *Spanish Amigo*TM, a bi-directional translation tool for English and Spanish which can be operated without leaving a Windows word processor, e.g. Microsoft Word for Windows, Microsoft Works for Windows, Microsoft Write, WordPerfect for Windows, Ami Pro for Windows, WordStar for Windows. The package, retailing at \$99.00, includes 150,000 words and phrases and the ability to add special terms and phrases to the user dictionary. It

requires IBM compatible 386 or greater, running Windows 3.0 or greater, in enhanced 386 mode, and a hard disk with at least 10MB of free disk space. The software joins the previous product from WestCliff *DosAmigos*. For information: WestCliff Software Inc., 343 Soqual Ave. Box 207, Santa Cruz, CA 95062; Tel: +1 (408) 459-8811; Fax: +1 (408) 459-8853.

New Software Release 7.5 from Logos Corporation

[Press release]

Logos Corporation of Mt. Arlington, NJ, announces a new release of its Multilingual Document Translation System at the American Translators' Association Conference.

This advanced Digital Text Translation software, release 7.5, is designed to translate a broad array of technical documentation, such as service and maintenance manuals, handbooks, technical correspondence and product descriptions where not only accurate transliteration is required, but also intelligent transfer of the meaning of sentences.

LOGOS release 7.5 is file format compatible with leading document publishing systems such as Interleaf, Frame, WordPerfect, Microsoft Word and others.

A notable new feature of release 7.5 is the Graphical User Interface (GUI), based upon OSF/Motif. This enhancement offers much easier handling of text and graphics in response to requests from LOGOS users and an increasingly DTP-literate marketplace.

Also new is the English-Italian language capability, which brings to seven the total number of language pairs available from LOGOS, as follows:

English-German	German-English
English-French	German-French
English-Italian	German-Italian
English-Spanish	

President and CEO of LOGOS, Jens Thomas Lück commented, "With the introduction of our Multilingual Document Translation System, release 7.5, LOGOS is reaching further into the world of standard-platform Digital Text Translation. Indeed, by using the Sun UNIX workstation as a server to a Local Area Network of X-terminals or PCs compatible with X-Window System Version II and OSF/Motif, a multi-user application can be developed for integrated multi-lingual electronic publishing. New LOGOS customers can be assured of receiving the full benefit from our 25 years of experience using computers to translate natural languages."

LOGOS also now offers custom translation consulting services which include: linguistic and text analysis, translation and terminology (dictionary and glossary building) services, and systems configuration and integration. Its highly qualified team of over thirty computational linguists and software engineers provides installation services, systems support, hotline support, and other information services.

A multipurpose linguistic processor under development in Moscow

Igor M. Boguslavsky, Leonid L. Iomdin

The Laboratory of Computational Linguistics of our Institute for Information Transmission Problems (Russian Academy of Sciences, Moscow), headed by Academician Jurij D. Apresjan, is developing a multipurpose linguistic processor (LP) designed to handle natural language texts. The linguistic processor can be viewed as a software system able to understand and produce natural language texts, that is, to emulate the language competence and performance of a human being. The creation of such a system is a fundamental problem of artificial intelligence. As an approach to this problem, a functional natural language model is built which simulates two major aspects of linguistic behaviour of man: understanding (analysis) of text and production (synthesis) of texts.

The model serves as a basis for two concrete software systems:

- (1) A high-quality machine translation system translating texts in information and computer science from English into Russian (ETAP);
- (2) A user-friendly interface for man-machine natural language communication.

Machine Translation System

In the ETAP MT system, the translation is fully automatic, i.e. it does not require human intervention at any stage of text processing. The current experimental operational version translates texts in the given subject domain at the level of syntactic structures of sentences represented as dependency trees whose nodes are labelled with names of lexemes of the processed sentence and arcs with names of syntactic dominance relations. Up to 50 syntactic relations are used, which provides very good control of sentence meaning. A large number of experiments have been carried out in which original texts from computer programming textbooks and reference manuals, as well as articles from recent issues of the *Computer* journal, were used. In addition, hundreds of experimental sentences were processed to check special syntactic constructions and translation rules. Sample translations obtained in these experiments are given below:

** Computer user interfaces have become more important with the increase in number of users and applications _ Komp'juterye interfejsy pol'zovatelja stali bolee vazhnymi s uvelicheniem v chisle pol'zovatelej i primenenij.*

** The personal-computer revolution and falling hardware prices made computers available to ever broader groups of people who use computers for a larger variety of tasks _ Perevorot personal'nogo komp'jutera i padajushchie ceny apparatnyx sredstv sdelali komp'jutery dostupnymi kak nikogda shirokim gruppam ljudej, kotorye ispol'zujut dlja bol'shego raznoobrazija zadach.*

** Initially, when computers were used by only a few people performing specialised tasks, it made some sense to require a high degree of user expertise _ Pervonachal'no, kgda komp'jutery byli ispol'zovany tol'ko neskol'kimi ljud'mi, vypolnjajushchimi specializirovannye zadachi, imelo smysl trebovat' vysokoj stepeni kvalifikacii pol'zovatelja.*

Currently, the system has the following parameters:

- completeness of morphological rules (analysis and generation): 100% [This means that the system can handle 100% of phenomena proper to (inflectional) morphology of the working languages.]
- completeness of syntactic analysis rules: about 85%
- completeness of transfer rules: about 85%
- completeness of syntactic synthesis rules: close to 90%
- morphological dictionary of English: 17,000 entries
- combinatorial English-Russian dictionary: 13,000 entries

Adequate and grammatically correct translations, easily understood by native speakers of the target language and comparable to high-quality human translations are averagely obtained for 60% of source language sentences. For the remaining 40% of sentences, rougher translations are offered (in some cases even word-for-word translations.)

We expect to substantially improve this parameter (up to 80% or more) in the nearest future due to (a) the expansion of the combinatorial dictionary, (b) through complex debugging of all linguistic components by testing the system on large text corpora, and (c) introduction of new technology of processing words unknown to the system, as the latter are currently responsible for the largest share of failures.

The system, implemented on a MicroVAX 3100 computer running under VMS operating system, takes 6 to 20 seconds per an average sentence 20 to 30 words long. All programs are written in C.

In addition to the main mode of operation which yields one (usually the most plausible) translation, a multiple translation mode is developed which offers alternative translations for syntactically and/or lexically ambiguous sentences. E.g., for the ambiguous sentence *They made a general remark that...* two different syntactic structures and respectively, two different translations are obtained: (a) *Oni sdelali obshchee zamechanie, chto...* (_ They made some common remark that...) and (b) *Oni vynudili generala otmetit', chto...* (_ They forced some general to remark that...) [The example is taken from real ETAP output.]

A technology is being developed which will make it possible to process whole texts rather than single sentences.

It is to be noted that the major part of linguistic knowledge used by the MT system is easily adaptable to the reverse translation option: from Russian to English. An experimental version of such a system has been successfully tested.

Natural Language Interface

The interface for man-machine communication provides a non-professional user with an effective means to address a database in virtually unrestricted natural language. The interface translates queries to relational type databases, formulated in an unconstrained language, into standard SQL expressions using full morphological, syntactic and semantic analysis. The interface can also be asked to paraphrase the query in order to make certain that it was properly understood. At a later stage the interface will be also equipped with a device capable of a reverse translation of the computer's responses into a natural language.

Foundation of Canadian Language Technology Institute

[From press release and publicity leaflet]

The Canadian Language Technology Institute (CLTI) was officially launched on 1st October 1993 on the campus of the Université de Moncton. The creation of this national leading-edge technology centre was made possible by a substantial contribution from the Federal government of Canada via a major research and development program in communication technology. Industry and Science Canada has given \$1.5 million and at the moment the Institute also benefits from support from the Université de Moncton, NB Tel, the University of New Brunswick, Treasury Board Canada and the Government of New Brunswick.

Jean-Guy Haché, president of the CLTI, said that the Institute's mission is the research, development and application of a technology that will one day allow computers to effectively use human language to execute such tasks as those involving the recognition of speech and hand-writing, as well as their reproduction and translation or interpretation. The CLTI will promote this technology to potential Canadian users and will ask them what their immediate or future needs are in that specific field in order to find practical solutions.

As the Canadian centre of excellence in the field, and based on the actual needs of its clients, CLTI will mobilize the expertise required to carry out world-class research, development and application activities, some 70% of which will be "operations oriented": transfer of technology; evaluating hardware/software; advocacy; innovation; and, acting as a clearing-house of technical information. The other 30% will be focused on education and research towards the development of better systems, in partnership with other centres domestically and internationally.

The present members of the Canadian Language Technology Institute, an autonomous nonprofit organization, are: the Université de Moncton, the University of New Brunswick, NB Tel, Assumption Life, IBM Canada, the Federal Translation Bureau, John Chandioix experts-conseils, Industry and Science Canada, and Economic Development and Tourism New Brunswick. The CLTI is supported by an international scientific advisory committee.

The goals of CLTI are:

- * To carry out world-class research, development and application work in language technology and related fields.

- * To serve as a technological incubator for Canadian researchers and practitioners of the private and public sectors with relevant projects.

- * To promote the research, innovation and application of language technology in Canada, and the use of linguistic tools derived therefrom.

- * To be the national information node in language technology and related fields, and make such information widely available.

- * To transfer relevant technologies, domestically or internationally, for Canadian research, development, application or marketing purposes.

* To develop expertise and provide professional development or training in language technology and related fields.

* To examine and apply to linguistics existing and emerging relevant technologies.

* To act as a catalyst/coordinator in mobilizing relevant Canadian efforts toward the achievement of meaningful and useful results.

* To advise appropriate public policy-makers on matters relating to language technology.

* To establish or advocate the establishment, as the case may be, of new businesses to market any marketable products/services developed by the Institute.

* To create appropriate consortia to address specific issues, needs or opportunities related to the Institute's areas of endeavour, and develop and implement solutions.

For further information contact: The Canadian Language Technology Institute, P.B.23130 Moncton, New Brunswick, Canada E1A 6S8. Tel: +1 (506) 855-2584; Fax: +1 (506) 855-4285; Email: jgh@clti-ictl.nb.ca

READERS' FORUM

This section of MT News International is for readers to express personal views of issues in the field of MT. The inclusion of items does not, of course, imply the endorsement or approval of the views expressed by either the editors or the International Association for Machine Translation.

The U.S. patent system and MT technology

Leighton Chong

As reported in a previous issue of MT News International, Systran asserted its U.S. Patent 4,702,212 for a computerized translation system against an MT competitor in mid-1993. The patent was issued in November 1987 in the name of Peter Toma, and was subsequently transferred to Systran. In reaction, the membership of IAMT passed a resolution condemning such use of patents and other intellectual property protections to inhibit MT development. In the opinion of this writer who is familiar with the intellectual property field, IAMT's condemnation of the intellectual property system is misplaced. The real problem is the anachronism of the U.S. patent laws.

Intellectual property protections have evolved to protect the creators of inventions and original works for limited periods of time so that some return can be realized for the genius and labor invested in their creations. Without an intellectual property system, there would be little incentive for persons to employ their creativity for useful ends since the fruits of their labor could always be stolen or copied by others. Traditional forms of intellectual property rights in the U.S. and most industrialized countries of the world include patents, trademarks, copyrights, and trade secrets.

Of the spectrum of intellectual property rights, patents are usually of greatest interest to companies involved in technology development because they confer the legal right to exclude all others from infringing the patented invention (innocently or intentionally) as well as the right to collect damages for past infringement. To safeguard against the grant of spurious patents, the U.S. Patent Office employs a highly qualified corps of technical examiners whose job is to examine patent applications against the available patent and technical literature and to ensure that the claims of issued patents define precisely only what the inventor has contributed that is new and would not have been obvious to others at the time the invention was made.

Europe, Japan, and many other industrialized countries have what is referred to as "first-to-file" patent systems. The first-to-file system typically recognizes the priority of applicants who file their patent applications first, thereby providing an incentive to file as soon as an invention has been

completed. The application is usually published in an official publication 18 months after filing in order to place others on notice of the claim of patent rights as soon as possible. If a patent is granted, the patent term is measured (typically 20 years) from the original filing date. The main advantages of the first-to-file system are the incentive for prompt filing, early published notice of a claim of patent rights, and fixing of the patent term to the application filing date. The main disadvantage is that large companies with the resources can file patent applications faster and in greater volume than small companies and individuals.

In contrast to most other countries of the world, the U.S. has what is referred to as a "first-to-invent" system. In this system, the priority of an inventor can be established by proving that the invention was made in the U.S. before other persons independently invented the same thing. Thus, there is less pressure for an applicant to file a patent application immediately, as long as the development of the invention has been well documented. After a U.S. patent application is filed, it is maintained in secrecy by the Patent Office until a patent is granted. Because of the many time-consuming procedures available to applicants to ensure that an application is fully considered before it is denied or granted, U.S. patent applications can, and often do, remain pending for several years. And the 17-year U.S. patent term is measured from the date of issuance, not the date of filing of the application.

In the case of the Systran patent, the original patent application was filed in March 1971, but the patent did not issue until November 1987. The pendency for 16 years was extraordinary, even for U.S. practice, and was caused primarily by an extended series of rejections of the patent claims by the U.S. examiner, the pursuit of an appeal through the Patent Office Board of Appeals and then the Federal Circuit Court of Appeals, and a lengthy lapse on a petition to withdraw the patent from issuance to allow a transferee of the patent rights time to reconsider the patent. The patentee was able to obtain allowance of the patent in the end only after severely limiting the scope of the claims of the patent.

The term of Systran's U.S. patent therefore extends until the year 2004, whereas in Japan, Europe, and other countries, the patent would have expired by 1991. Thus, the outrage over a patent that extends into the next century, based upon technology that was developed in the 1960's, should be directed at the anachronism of U.S. patent laws, rather than the inhibiting effect that recognition of intellectual property rights might have on technology development. To the contrary, technology development would certainly be inhibited by rampant secrecy if a system of intellectual property protection were not in place.

U.S. patent rules do not discourage an applicant from undertaking procedures that can cause long-delayed issuance. Discussions have been going on for almost 40 years to change U.S. patent laws so that they can be more in harmony with the rest of the world. However, the U.S. view is that its patent system favors the rights of first inventors, rather than those first to file, and puts small companies and individual inventors on an equal footing with large companies. Therefore, there has been very little mandate within the U.S. to change its patent laws to a first-to-file system.

The U.S. may be forced to change its present patent system if GATT negotiations move forward toward adoption of new multilateral trade rules. Under the proposed new GATT rules, the U.S. would be required to extend the protections of its first-to-invent system to foreign applicants based upon their acts of invention in foreign countries or else abandon it. However, it may be many years before the adoption by participating countries of new GATT rules, much less change of the U.S. patent laws.

In the meantime, the MT industry in the U.S. can take several different approaches to dealing with the Systran patent. One prudent step for MT companies which use a system of machine translation similar to that described in the Systran patent should be to obtain the opinion of competent patent counsel whether they might be infringing the patent claims. The patent claims were severely limited in order to obtain their issuance, so other MT systems may actually be non-infringing even though they may use a similar technical approach. For example, the patent file history shows that each of the main Claims 1, 15, 18, 19, and 22 of the patent had the significant limitations recited in the last paragraph(s) thereof in order to qualify the more general translation

system described in the preceding paragraphs. A company would be entitled to rely upon an opinion of non-infringement by competent patent counsel to limit its liability to compensatory damages in the event a court might later determine that the patent is infringed. If infringement is a significant possibility, a company could negotiate for a license under the Systran patent.

Any individual, company, or association can also request (for a relatively modest government fee of \$2,250) that the U.S. Patent Office re-examine the Systran patent to determine whether it is valid over any publications or public uses disclosing the translation technology described in the patent prior to 1970 which was not known to the examiner on the original patent application. Although individual companies charged with patent infringement often prefer to litigate the validity of a patent in a court setting because of procedural safeguards, re-examination of a patent by technically qualified U.S. examiners (rather than by a jury or judge) could settle the question of the validity of the Systran patent for the industry as a whole in a single proceeding that is more expert, consistent, quicker, and far cheaper than court litigation. As a recent example, one industry association put out a call on Internet for references that might be used to contest the validity of a patent and received hundreds of responses. This approach can save thousands of dollars in search costs and can result in finding pertinent prior art references to bring to the attention of the Patent Office in re-examination.

As a fledgling industry, MT companies should also familiarize themselves with the basic framework of the patent system and intellectual property protections not only in the U.S. but also other countries of the world as well. MT technology has applicability throughout the world. A small company that has developed new MT technology might find that the intellectual property system is vital to establishing a competitive position in global markets.

Leighton Chong is founder and president of Trans-Link International Corporation, a company engaged in research and development of machine translation systems in Honolulu, Hawaii. He is a patentee of U.S. Patent 5,175,684 for auto-loading of dictionaries of an MT system in a telecommunications environment. He is also a patent attorney in New York City and senior partner of the intellectual property firm of Ostrager, Chong & Flaherty.

How Lingvistica '93 MT Systems are Used (1991-1993)

Michael Blekhman

We [Lingvistica '93 Co.] are offering three machine translation systems, having the same design features and looking quite similar:

PARS - for English-Russian and Russian-English translation;

RUMP - for Russian-Ukrainian and Ukrainian-Russian

PARSU - for English-Ukrainian-English.

1. The Ukrainian-oriented systems were developed to meet the demands of the rather complicated language situation in Ukrainian. The problem is that most people in this country are native speakers of Russian, though most of them understand simple Ukrainian. West Ukraine is but the only region where Ukrainian is spoken "since napkins", though Russian is also spoken there.

The situation with Ukrainian, which, by the way, is a charming language, with a mighty literature, resembles that with Hebrew. The latter was revived as a state language, which is to be done now as to Ukrainian. According to the Law on the Languages, adopted 3 years ago, Ukrainian is to be the official language in the country, which, as I understand, demands serious efforts in computer technology.

Speaking of the RUMP system, it was developed in 1990 to meet the demands. Since then, the system has been implemented at dozens of enterprises, on hundreds of computers across the country. It comprises a two-way Russian-Ukrainian dictionary of political terms and general usage words and phrases, though, as in all our systems, the dictionary can be and is enlarged, both by me and by the end users. It is mostly used to translate official documents at political establishments, such as the Parliament, procurator's offices, various governmental bodies, etc. In 80-90% of

applications, it is used to translate from Russian into Ukrainian; however, with creation of Ukrainian-language data bases of official documents, it is also applied to translate from Ukrainian into Russian, so the proportion will be gradually changing, to meet the 50/50 level in future.

At present, I am developing a special dictionary to translate police information which comes on-line from the local stations to the Kharkov Central Police Department. The problem for the end users is the same: the cops only speak Russian, while the boss (the General) says the law is to be followed, so the information that comes to him must be in Ukrainian. This particular situation seems a bit artificial to me: I believe that the principal thing for the police is to catch criminals rather than study languages. However, RUMP will be helpful, and this is the most important thing for me, as a linguist.

As to really useful applications, I know of several cases when RUMP was used to translate books from Russian into Ukrainian. At the Palace of Children in Kiev, four books on computers were machine translated into Ukrainian this year, with 80-90% accuracy. They came from Russia on floppy disks. In the process of translation, terms on computers were entered into the system dictionary.

I, personally, had my book [on computational linguistics] machine translated into Ukrainian. Though Ukrainian is my second native tongue, the book was originally written in Russian. The machine product required extensive post-editing because the book comprised a lot of 'new' terms and meanings, and the syntax was too complicated in some cases, which convinced me in the necessity of improving the transfer procedure. However, I perceived that, without access to an MT system, I would have never forced myself to translate 0.5 megabytes manually.

2. The PARS system is broadly used both in Ukraine and in other ex-Union countries, such as Russia, Kazakhstan, Georgia. The geography is rather wide, from Lvov in the West to Sakhalin in the Far East, in Russia. While RUMP is so far mostly used at official structures, PARS has quite a number of individual end users. Among the latter, I see two 'subgroups': professional translators and Russian speaking scientists and engineers. A very interesting tendency is partial reorientation of the users from English-Russian to Russian-English translation, which is caused by Russian and Ukrainian penetrating Western markets. What is very pleasant for me, some users need both English-Russian and Russian-English translation, which proves the initial 'two-way' philosophy was correct.

In general, the situation is far from being cloudless, though quite a number of people thank me for having designed a useful translation instrument.

There are people who have negative attitude to the PARS system, saying that it is easier to translate the whole text manually than to post-edit the machine product. The most unexpected thing, however, is that the higher professional skills, the better attitude to MT. I discussed the problem with a lady translator who worked at the Izvestia Concern, Moscow, where PARS is implemented. I was sitting there for about an hour and a half, working with PARS, while she was manually translating VWD information, using a 2-window text editor. I noticed that she only managed to translate two documents, within that time, though, certainly, she was supposed to give 'ideal' translation, which is unattainable for any software. It only took PARS a minute to rough translate one document of average size on an IBM/286 PC, and the translation quality seemed not bad to me, the contents being quite clear. However, when asked if PARS could help her, the translator said, "You see, it seems useless since the machine product needs too much post-editing, and I only know three keys on the computer keyboard, which makes post-editing a terribly boring task for me."

On the contrary, one of my clients, Vladimir Kolykhmatov, who is a very experienced translator working for the Moscow Agency of the Dupont Company, translating 80-90% from Russian into English and 10-20% from English into Russian, says he cannot work without PARS, saying "I am too lazy to do the job all on my own; I prefer to post-edit the raw translation instead of typing in everything with my fingers." When shown the translations his lady colleague had rejected, he evaluated them rather high, saying "I don't understand what on earth those guys want!"

I am very glad PARS has also been appreciated by non-translator end-users. In particular, Dr Boris Piskunov of the South Sakhalin Institute of Geology sent me a PARS-made English translation

of his scientific report he was supposed to submit to his South Korean partners, with the commentary that he was satisfied as the volume of post-editing needed was not large. I have just received similar information from Dr Vladimir Petushkov of Ukrainian Welding Institute. Both of them have mastered entering new words into PARS dictionaries, saying, however, that the dictionary extending routine needs improving since it is not easy to code Russian words, especially the verbs. This is really so, although the program gives a lot of Helps. That is why I decided to work out a much more flexible routine, with automatic coding. It is under way now.

[For technical details of the systems mentioned see the section in this issue on 'Products and Systems'.]

PUBLICATIONS ANNOUNCED AND RECEIVED

Natural Language Software Registry Second Edition Now Available

The Registry team is proud to announce the 2nd edition of the Natural Language Software Registry. Its purpose is to facilitate the exchange and evaluation of natural language processing software within the research community. To this end, the Natural Language Software Registry is cataloguing natural language software projects, both commercial and non-commercial. The new updated and enlarged version contains more than 100 descriptions of natural processing software. Registry listings include: speech signal processors, morphological analyzers, parsers, semantic and pragmatic analyzer, generation programs, knowledge representation systems, multicomponent systems, NLP-Tools, applications programs.

We are most thankful to these and other research teams who have taken the time to document their achievements. 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 questionnaire available from the sources below.

In the coming months, the Registry team will also be involved in a study of linguistic software repositories for the European Community. Researchers who regard their systems as potential EC resources are encouraged to contact us for consideration of their software in this pilot study.

The document and the questionnaire can be obtained via ftp from Germany: ftp.dfki.uni-sb.de (134.96.188.252), directory: pub/registry, password: anonymous; by email from: registry@dfki.uni-sb.de; and by post from: Natural Language Software Registry, Deutsches Forschungsinstitut für Künstliche Intelligenz (DFKI), Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany

The Registry is also available via ftp in the US from the Center for Lexical Research: crlftp.nmsu.edu (128.123.1.33) [The directory is pub/non-lexical/NL_Software_Registry], and from the Design Research Institute: dri.cornell.edu (128.84.180.39) [The directory is /pub/Natural_Language_Software_Registry/ or /pub/NLSR]

Saarbrücken Report on MT in Translator Training Courses

On 31 March 1993 a five-year project ended at Saarbrücken University on integrating natural language processing into diploma courses for translators and interpreters. [For bibliographic details of the report see 'Publications Received'.] The course components cover the theoretical foundations of language processing, including general data processing for the particular needs of translators and interpreters; foundations of terminology, computer-based terminology work and machine and machine-aided translation; applications, functions and performance of systems for text processing, for terminology management and retrieval, for machine (aided) translation, and the practicalities of working with these systems; and the pursuit and evaluation of basic and applied research in seminar

and diploma projects. In the MT seminars demonstrations of current commercial or research systems have been given, e.g. SUSY, Systran, Logos, Eurotra, Globalink, Translator's Workbench II, Transit, TranslationManager/2. Additional seminars take place which explore aspects more deeply, e.g. foundations and problems of machine translation; grammatical models applied in language analysis; new technologies for the translator workstation; introduction to terminography; linguistic models in NLP; machine translation systems; terminology databanks; theory and practice of computer-based terminology work; etc. The project report details the structure of the seminars and courses, describes the adoption of the 'model' at other universities, and looks at future prospects. [A summary of the report is given by Karl-Heinz Freigang and Uwe Reinke in *Lebende Sprachen* 3/1993, 100-104.]

Quebec Report on the Language Industries

In July 1993, an interministerial group in the Quebec Province reported on the prospects for the language industries [for bibliographic details see 'Publications Received'.] The group was set up in Spring 1992 comprising representatives from the ministries of Communications, of Industry, of Commerce and Technology, and of International affairs, from the Office de la Langue Française, the Centre Francophone de Recherche en Informatique des Organisations (CEFRO), and from the Observatoire Québécois des Industries de la Langue. The working document presents the situation in Quebec, in Canada and the world, describes the economic, technological, cultural and linguistic issues, attempts to define the market for the language industries in Quebec and outlines the conditions necessary for their development. As a result of its investigations the group has now set up four working committees to explore in depth: the development of human resources; the markets; research and development; and the international dimension.

New Introductory Text for MT

In December 1993, NCC Blackwell (Manchester/Oxford) published *Machine translation, an introductory guide*, written by members of Essex University (Doug Arnold, Lorna Balkan, R.Lee Humphreys, Siety Meijer and Louisa Sadler) - for full details see 'Publications Received'. The authors have aimed "to write a book which would be genuinely introductory (in the sense of not presupposing a background in any relevant discipline), but which would look at all aspects of Machine Translation." In an informal colloquial (British) style, they dispose first of popular misconceptions about MT and give an idea of how MT systems are used in practice, then provide a simple introduction to the basic linguistic problems, describe the major approaches (direct, transfer, and interlingua), deal with requirements for dictionaries in MT and how they differ from paper dictionaries, sketch out the problems of meaning (lexical, pragmatic and real world semantics), explain controlled and sublanguage systems, outline the basics of evaluation, and conclude with a brief survey of current approaches (constraint-based, example-based and statistical MT). The authors provide a glossary of terms together with useful addresses and guidance to books and articles for further reading. Technical detail is minimized for the sake of comprehensibility, and the book could well serve as a good 'taster' of MT for those unfamiliar with the basic issues and problems of the field.

TEXT Technology

[From LINGUIST List: Vol-4-1051]

Due to the unanticipated level of its success, the journal *TEXT Technology* will be substantially expanded. Starting, with Volume 4 (the 1994 calendar year), each issue will contain more articles and reviews -- many of them presenting subjects in greater depth. The journal will be published quarterly, and the Editorial Board will be expanded. The format will also change to 7-by-9-inch pages with perfect binding.

The full title of the publication will now more completely describe its contents: *TEXT Technology - The Journal of Computer Text Processing*.

TEXT Technology will continue to publish articles and reviews about all facets of using computers for the creation, processing, communication, and analysis of texts. It is designed for academic and corporate researchers, writers, editors, and teachers. The quarterly journal contains timely reviews of books and software, discussions of applications for the analysis of literary works and other texts, bibliographic citations, and much more.

Recent issues of *TEXT Technology* have contained articles about the ideal computing lab for composition classes, counting the amount of quotation in novels, programming in Icon, converting documents from Macintosh to PC formats, as well as reviews of books about Internet and reviews of new versions of WordPerfect, AmiPro, and OS/2.

Submissions of articles are welcome. They should be sent to the Editor as ASCII files via email to JohnsonE@columbia.dsu.edu. Writers of book or software reviews are encouraged to contact the Editor before submitting reviews. Authors will normally receive notices of acceptance and referees' comments promptly via email.

New yearly subscription rates are in effect immediately: in the U.S., Individuals: \$45.00; Institutions: \$72.00. Canadian orders add \$7.00; all other nations add \$15.00 (all prices U.S. funds).

To subscribe using a MasterCard or Visa credit card, send name and address, card number and expiration date via email to LangnerS@columbia.dsu.edu.

To subscribe by regular mail, send credit card information, check, or institutional purchase order to TEXT Technology, 114 Beadle Hall, Dakota State University, SD 57042-1799 USA.

Arabic Computing

[Publicity leaflet]

Taking over from where *Processing Arabic* left off, Arabic Computing will offer the latest work and thinking connected with all aspects of the use of computer techniques in relating to the Arabic language. It will take in issues from the areas of natural language parsing and generation, artificial intelligence applications, computer assisted instruction and language learning, cataloguing, compilation of corpora, data storage and retrieval, machine translation, optical scanning, statistics and text compression as well as research and activities in a number of other domains.

Arabic Computing is supported by the International Association of Arabic Computing (IAAC), a mondial workgroup of people who as developers and end-users are interested in the use of computer techniques in relation to the Arabic language. Papers should be sent as hard copy and either on disk or via email to: Dr Everhard Ditters, General Editor Arabic Computing, TCMO, Nijmegen University, PO Box 9103, 6500 HD Nijmegen, The Netherlands. Email: u279000@hnykun11.bitnet.

Arabic Computing will appear three times a year, the first issue in January 1994. Personal subscriptions £30.00; institutional subscriptions £70.00. Orders to: Intellect, 108-110 London Road, Headington, Oxford OX3 9AW, U.K.

PUBLICATIONS RECEIVED

Journals

AAMT Journal: Journal of the Asia-Pacific Association for Machine Translation, *no.4 (September 1993)*. p.1: New dream for machine translation (Atsushi Asada). -- p.2-7: 3rd General Members' Assembly at kyobashi-kaikan Hall in Ginza, June 25, 1993. -- p.8-10: MT Summit IV ended successfully in Kobe. -- p.11: Machine translation research in Indonesia (Darmawan Sukmadjaja). -- p.12: University education and MT (Masaru Tomita). -- p.13: Survey reports from three work groups [of AAMT]. -- p.14-15: Translation

workbench PIVOT/JE R5.1 from NEC. -- p.16-17: Nova introduced PC-Transer/je. -- p.18: Conditional expectations (Takamasa Imai). -- p.19, 23: MT Utilization Technic W.G., extract from "Survey report" [translated by NEC's PIVOT system]. -- p.20-22: Computer environments for machine translation use (Susumu Akamine).

Computational Linguistics, *vol.19, no.2 (June 1993) - Special issue on Using Large Corpora II*: p.219-241: Using register-diversified corpora for general language studies (Douglas Biber). -- p.243-262: From grammar to lexicon: unsupervised learning of lexical syntax (Michael R.Brent). -- p.263-311: The mathematics of statistical machine translation: parameter estimation (Peter F.Brown, Stephen A.Della Pietra, Vincent J.Della Pietra, and Robert L.Mercer). -- p.313-330: Building a large annotated corpus of English: the Penn Treebank (Mitchell P.Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz). --p.331-358: Lexical semantic techniques for corpus analysis (James Pustejovsky, Sabine Bergler, and Peter Anick). -- p.359-382: Coping with ambiguity and unknown words through probabilistic models (Ralph Weischedel, Marie Meteer, Richard Schwartz, Lance Ramshaw, and Jeff Palmucci). -- p.383-400: Book reviews.

Language Industry Monitor, *no.17 (September-October 1993)*. p.1-4: Focus on Japan: why there might be a Japanese MT system in your future. -- p.4-6: MT Summit IV [report of conference]. -- p.7-9: A translation factory [on IBS use of NEC's PIVOT system.] -- p.9-10: Fujitsu [on ATLAS]. -- p.10-11: Toshiba [research on ASTRANSAC]. -- p.11-12: CSK [development of Argo]. -- p.12: MT Labs, Inc. [owners of Bravice software]. -- p.13: NHK [use of Catena STAR system]. -- p.13-14: TMI '93 notes [mainly on Microsoft and CMU presentations].

no.18 (November-December 1993). p.6: ISSCO: neutral but not passive. -- p.9-10: Cap Volmac's tailor-made MT systems. -- p.11-12: A heavy METAL user [Union Bank of Switzerland]. -- p.14: Canadian Language Technology Institute.

Language International, *vol.5 no.4 (August 1993)*. p.9: Systran now on PS/2. -- p.10: New pre-edit system (Toon Witkam) [Pre-translation system from Brussels-based journalist Paul Ruëll].

vol.5 no.5 (October 1993). p.5-6: Evaluating machine translation in the 1990s, Machine Translation Summit IV, Kobe, Japan, 20-22 July 1993 (Bertie Kaal).

Literary & Linguistic Computing, *vol.8 no.3, 1993*. p.113-130: Design and implementation of a spelling checker for Turkish (Aysin Solak and Kemal Oflazer).

Machine Translation, *vol.8, no.1-2, 1993: Special issue on evaluation of MT systems*. p.1-24: Evaluation: an assessment (Doug Arnold, Louisa Sadler, and R.Lee Humphreys). -- p.25-27: Evaluation of MT systems at Union Bank of Switzerland (Doris Albisser). -- p.29-38: Automatic test suite generation (Doug Arnold, Dave Moffat, Louisa Sadler, and Andrew Way). -- p.39-47: Module-level testing for natural language understanding (Sharon Flank, Aaron Temin, Hatte Blejer, Andrew Kehler, and Sherman Greenstein). -- p.49-58: A first-pass approach for evaluating machine translation systems (Pamela W.Jordan, Bonnie J.Dorr, and John W.Benoit). -- p.59-66: Evaluation of MT systems: a programmatic view (Steven Krauwer). -- p.67-75: Constructive machine translation evaluation (Stephen Minnis). -- p.77-84: Benchmark investigation/identification project (Jeannette G.Neal, Elissa L.Feit, and Christine Montgomery). -- p.85-107: A diagnostic tool for German syntax (John Nerbonne, Klaus Netter, Abdel Kader Diagne, Judith Klein, and Ludwig Dickmann). -- p.109-116: A procedure for the evaluation and improvement of an MT system by the end-user (Brigitte Roudaud, Maria Claudia Puerta, and Otemea Gamrat). -- p.117-126: Automatic evaluation of output quality for machine translation systems (Yu Shiwen).

vol.8, no.3, 1993. p.127-145: Automatically creating lexical entries for ULTRA, a multilingual system (David Farwell, Louisa Guthrie, and Yorick Wilks). -- p.147-173: What can be learned from raw texts? (Roberto Basili, Maria Teresa Pazienza and Paola Velardi). -- p.175-201: Acquisition of selectional patterns in sublanguages (Roberto Basili, Maria Teresa Pazienza and Paola Velardi). p.203-207: Symposium report: AAAI 1993 Spring Symposium on Building Lexicons for Machine Translation (Clare R.Voss).

Books

Bouillon, Pierrette, and Clas, André (eds.): *La traductique: études et recherches de traduction par ordinateur*. Montréal: Presses de l'Université de Montréal, 1993. (Universités Francophones AUPELF/UREF) 507 pp. ISBN: 2-7606-1616-9. FF 170.

Contents: p.13-20: Introduction et bref historique (Pierrette Bouillon). -- p.23-26: Les problèmes posés par la traduction automatique (Jean-François Hùe). -- p.27-34: Préalables linguistiques à la construction d'un système de traduction automatique (Guy Bourquin). -- p.35-40: La typologie des textes: son importance pour la traduction automatique (Alan Melby). --p.43-63: Classification générale des systèmes de traduction automatique (Laurence Jacqmin). -- p.64-76: Sur la conception du transfert (Doug Arnold). -- p.77-90: Divergences de traduction et architectures de transfert (Françoise Vandooren). -- p.91-108: L'interlangue et le traitement du sens dans les systèmes de traduction automatique (Sergei Nirenburg). -- p.109-148: La TAO comme technologie scientifique: le cas de la traduction automatique fondée sur le dialogue (Christian Boitet). -- p.149-166: La traduction automatique basée sur l'exemple ou sur les corpus (Harold L.Somers). -- p.167-196: Le lexique: quelques problèmes de description et de représentation lexicale pour la traduction automatique (Ulrich Heid). -- p.199-211: Théorie linguistique et traitement automatique du langage (Christopher Laenzlinger et Éric Wehrli). -- p.212-227: Grammaires d'unification et traduction automatique (Dominique Estival). --p.228-240: Deux théories adaptées à la traduction automatique: GPSG et HPSG (Philippe Blache). -- p.241-257: Sémantique et traduction automatique (Graham Russell). -- p.261-269: Sur l'évaluation des systèmes de traduction assistés par ordinateur (Margaret King). -- p.270-277: Acheter un système de traduction automatique aujourd'hui (Kirsten Falkedal). -- p.281-287: Le poste de travail du traducteur (PTT), ou les aides à la traduction (Elliott Macklovitch). -- p.288-306: Les corpus bilingues: une nouvelle ressource pour le traducteur (Pierre Isabelle et Susan Warwick-Armstrong). -- p.307-321: Traduction automatique et programmation en logique (Patrick Saint-Dizier). -- p.325-333: L'état de recherches en traduction automatique au CELTA (Arlette Attali, et al.). -- p.334-348: Questions de traduction automatique au Computing Research Laboratory (CRL) (Stephen Heimreich, et al.). -- p.349-363: La segmentation du chinois - une étape cruciale pour la traduction automatique du chinois (Wanying Jin et Jian-Yun Nie). -- p.364-376: L'unité lexicale dans le système de traduction assistée par ordinateur METAL (Thierry Fontenelle, et al.). -- p.377-391: Le système de traduction automatique ETAP (Jurij D.Apresjan, et al.). -- p.392-407: Discours et traduction automatique: une approche interlangue basée sur les connaissances (Christine Defrise). -- p.408-422: La traduction de la langue parlée (Hiroaki Kitano). -- p.423-432: Vers un système de traduction interactif (Éric Wehrli). -- p.433-445: Quelques notions de stylistique pour le système de traduction des bulletins d'avalanches (Pierrette Bouillon et Katharina Boesefeldt). -- p.446-453: Réseaux sémantiques et génération (Lisbeth Degand, et al.). -- p.454-468: Analyse multilingue et représentation du contenu informationnel des textes médicaux (Anne-Marie Rassinoux, et al.). -- p.469-499: Bibliographie générale.

Arnold, D., Balkan, L., Humphreys, R.Lee, Meijer, S. and Sadler, L.: *Machine translation, an introductory guide*. Manchester/ Oxford: NCC/Blackwell, 1994 [c.1993]. viii, 240 pp. ISBN: 1-85554-246-3 (Hardback) £ 35.00; 1-85554-217-x (Paperback) £18.99. **Contents:** Introduction and overview. -- Machine translation in practice. -- Representation and processing. --Machine translation engines. -- Dictionaries. -- Translation problems. -- Representation and processing revisited: meaning. -- Input. -- Evaluating MT systems. -- New directions in MT. --Glossary. -- Bibliography. -- Index.

Conference proceedings

L'environnement traductionnel: la station de travail du traducteur de l'an 2001. Journées scientifiques du Réseau thématique de recherche "Lexicologie, terminologie, traduction", Mons 25-27 avril 1991. [Ed.] André Clas, Hayssam Safar. Sillery (Quebec): Presses de l'Université de Québec, 1992. xxi,374 pp. ISBN: 2-7605-0705-x. FF 190, Can\$ 35. **Contents include:** p.3-20: Quelle automatisations de la traduction peut-on souhaiter et réaliser sur les stations de travail individuelles? (C.Boitet). -- p.21-30: L'environnement sémantique, aide de transfert lexical en TA (A.Attali, A.Euvrard). -- p.31-48: Problèmes de désambiguïsation interactive en TAO personnelle (H.Blanchon). --p.51-58: Le module lexical dans un système de traduction automatique (L.Degand, G.Everaert, J.Lerot). -- p.59-68: Pour une approche interactive au problème de la traduction automatique (E.Wehrli). -- p.69-78: La traduction automatique des bulletins d'avalanches de la Suisse (P.Bouillon, K.Boesefeldt). -- p.79-94: Le projet EUROTRA (C.Jadelot, D.Boussel). -- p.95-108: ODILE-2, un outil pour traducteurs occasionnels sur Macintosh (M.Lafourcade). -- p.109-134: Intégration des connaissances en génie linguistique (J.Y.Morin). -- p.135-138: Poste de travail de traduction par l'informatisation d'un dictionnaire français-arabe (S.Taleb). -- p.139-150: Traduction et terminologie (D.Bachut, R.Gerber). -- p.151-154: Pour le traducteur: un poste de travail à trois niveaux d'assistance (A.Melby). -- p.155-160: Le poste de travail de traducteur à l'ONU - horizon 2001 (M.Gainet). -- p.161-170: La technologie au service du traducteur (K.Tremblay). -- p.171-182: PTT-2: la deuxième version du poste de travail du traducteur du CCRIT (E.Macklovitch). -- p.183-188: Trois scénarios possibles pour la station de

travail du traducteur de l'an 2001 (P.Sören Kjaersgaard). -- p.189-200: Un projet de dictionnaire intelligent (J.M.Maes). -- p.201-206: EURODICAUTOM, ou la terminologie en l'an 2001 (A.Reichling). --p.207-212: La modernisation de la traduction à la Commission et ses moyens (K.Van der Horst)

Machine translation today: Translating and the Computer 15. Papers presented at a conference jointly sponsored by Aslib, The Association for Information Management, Aslib Technical Translation Group, Institute of Translation and Interpreting, European Association for Machine Translation, 18-19 November 1993. London: Aslib, 1993. 196 pp. **Contents include:** p.1-17: Developments in machine translation research in the US (Yorick Wilks). -- p.19-22: Machine translation, myth, muddle or reality (Seamus Derrington). -- p.35-39: Controlled English with or without machine translation (Arthur Lee). -- p.83-98: New trends in terminology processing and implications for practical translation (Blaise Nkwenti-Azeh). -- p.111-120: Evaluation of machine translation software and methods (Margaret King). --p.121-130: Forty ways to skin a cat: users report on machine translation (Veronica Lawson and Muriel Vasconcellos). -- p.181-192: Future translation workbenches: some essential requirements (Richard Birch).

Reports

Angelova, Galia, and Winschiers, Heike: *The role of knowledge bases in machine aided translation.* Project 'German-Bulgarian Machine-Aided Translation', Technical Report 2-93. [Hamburg: Dept. Computer Science, Univ.Hamburg], July 1993.

Fischer, Inge; Freigang, Karl-Heinz; Mayer, Felix; and Reinke, Uwe: *Modellversuch Studienkomponente Sprachdatenverarbeitung in der Übersetzer- und Dolmetscherausbildung.* Abschlussbericht (1.4.1988 - 31.3.1993), Saarbrücken, März 1993; [and] Skripten des Saarbrücker Modellversuchs, Saarbrücken, März 1993.

Groupe de Travail Interministériel sur les Industries de la Langue (Québec, Canada): *Les industries de la langue: enjeux et perspectives.* [Québec], juillet 1993.

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.

AMTA Yellow Book Now Available

AMTA is pleased to announce publication of the Yellow Book, a handy locator of AMTA members and others who are involved in, or interested in, machine translation. This is a "first approximation"; next year it should be greatly expanded, possibly with the full membership lists of AAMT and EAMT as well.

Copies are available from AMTA Headquarters and may be requested using the form below. The price is \$6.00 plus \$4.00 for postage and handling.