MT News International

Newsletter of the International Association for Machine Translation

In this issue . . .

Of special note . . .

- **MT Summit Reports** *(page 14)*

- **Report on Data-driven MT Workshop** *(page 9)*

- **AMTA 2002 Call for Papers** *(page 14)*

- **Profile: Bente Maegaard** *(page 12)*

- **SDL International's Hedley Rees-Evans, and fresh perspectives from young researchers** *(page 10)*

Spotlight on the News

Apptek Founders Buy Back Company

Applications Technology, also known as Apptek, was one of the many language technology companies acquired by Lernout & Hauspie in the late 1990s. Apptek, an MT software developer specializing in Arabic is located in McLean, Virginia, and has remained a wholly-owned subsidiary of L&H. Mudar Yaghi and Mohammad Shihadah, who were partial owners of Applications Technology, now seek to buy the company back in the course of L&H's liquidation. The two parties have agreed to a price of approximately US$2 million. L&H had purchased Apptek in 1998 for US$17.5 million.

An auction the week of November 26th sought to complete liquidation of the remaining assets of Lernout & Hauspie, now in the final stages of bankruptcy proceedings in both the U. S. and Belgium. While the outcome of the auction has been contested, all of the machine translation units of L&H have been successfully sold to new

The End of Logos Corporation

MTNI has been unsuccessful in getting news or statements from the German language services provider global words GmbH concerning the Logos MT system. An extremely brief announcement appeared earlier this year stating that global words was "merging with" and "taking over" Logos Corporation. Bud Scott, founder of Logos, posted the information below to the MT-List on November 6th. We hope that more information will be forthcoming!

It will be of interest to the MT community to know that one of the oldest, pioneering companies in this field, Logos Corporation, the company that developed the Logos System, has recently ceased to exist as an operational entity. In the early 1970's, Logos Corporation built an English-Vietnamese production MT system under contract to the U.S. Defense Department. The system had translated thousands of pages of military manuals by the time the conflict ended. The then director of the Defense

Apptek

owners. The technologies originally developed by Globalink and Neocor and purchased by L&H in 1997 and 1998 were made part of L&H's Mendez translation unit which was sold to Bowne Global Solutions in August, 2001. The Metal machine translation system, bought by L&H in 1997 from GMS became part of SAIL Labs, which spun off from L&H in 1999. ❏

Logos

Department Research and Engineering (DDR&E), Dr. John Forster, stated that the Logos E-V System had established for the first time the feasibility of large-scale machine translation (DDR&E Annual Report, 1973). This was the first positive official assessment since AL-PAC trashed the notion of MT in 1966.

Logos Corporation was founded in 1969 by members of a Catholic kibbutz living on a large farm in the foothills of the Catskills in New York State. Most of the staff in the company's earlier years were drawn from that community. Anyone interested in knowing more about the company's founding and early history can obtain The Logos Story, by emailing the undersigned.

In its heyday, Logos Corporation had over 100 linguists and computer people working on the system, with development operations in New York and New Jersey, Boston, Santa Clara, Saigon, Teheran, Salerno, Montreal, Frankfurt and Saarbruecken. The system is considered by some experts as uniquely principled, with output quality in certain domains ranking among the very best. The system's theoretical motivations are described in a paper at http://www.iai.uni-sb.de/~carl/iaiwp/p11/index.html

It is expected that the Logos System, which has specialized in corporate document translation (from English and German), will continue to be marketed and developed by a successor organization.

Bernard (Bud) Scott, Founder, Logos Corporation Chief Architect, Logos System bscott22@verizon.net ❏

Professor Nagao Re-elected as President of Kyoto University

On August 29th, Kyoto University re-elected Professor Makoto Nagao. The term of office begins December 16, and lasts for 2 years. The president is elected by all academic personnel and serves four years in the first term, and for a further two years if re-elected. Makoto Nagao was elected the 23rd President of Kyoto University in December 1997. Dr. Nagao was the

Dean of the Graduate School of Engineering at the time of his appointment as the President. He previously served as the Director of the University Library (1995-1997), a member of the University Council (1995-1997,1992-1994). He also was the Director of the Data Processing Center (1986-1990).

In addition to his role at Kyoto University, he assumed the presidency of the Japan Association of National Universities in April 2001. Dr. Nagao was awarded the IEEE Emanuel R. Piore Award in 1993, Purple Ribbon Medal from the Japanese Government in 1997, an honorary Doctor of Science degree from the University of Nottingham in 1999 and many other prizes. ❏

Lionbridge and Sail Labs Form Alliance

[Adapted from press release]

With this development, 4 of the top localization companies now own, or are strongly allied with suppliers of, machine translation (the others are Bowne Global Solutions, SDL International and ALPNET. Berlitz GlobalNET has a loose alliance with SYSTRAN). This is an impressive step forward for MT, which was formerly dismissed as unsuitable for localization. —ed

Lionbridge Technologies, Inc. (Nasdaq: LIOX), a provider of solutions for worldwide deployment of technology and content, and Sail Labs, a provider of technology for cross-lingual information processing and multilingual information exploration, have entered into an alliance to jointly deliver enterprise language management solutions. The two companies have successfully combined translation memory, machine translation (MT), terminology management and workflow automation for a large enterprise software organization, and will now jointly develop and market customized implementations for other enterprises requiring fast, affordable, high-quality translation.

"For effective content globalization, customers need to integrate machine translation with other enterprise language management functions," said Jorgen Danielsen, Lionbridge vice president of product development. "By combining the Lionbridge Globalization Platform components for connectivity, translation memory and workflow with the Sail Labs MT engines, we can provide our clients with comprehensive solutions for high-volume requirements."

For example, one Lionbridge eSupport implementation requires processing more than 100,000 words per day, using translation professionals to rapidly post-edit the result. With a combined MT and translation memory approach using solutions from Lionbridge and Sail Labs, urgent, multilingual, publication-grade content is produced in less than four hours.

"As seen from Lionbridge's long-standing client relationships, they have a proven track record of implementing language technology in complex, global environments and understand the importance of impeccable project management and customer service for long-term customer satisfaction," said Jurgen Werny, Sail Labs senior vice president and COO. "Working together, Lionbridge and Sail Labs have delivered some of the most advanced language management solutions available. We look forward to expanding this relationship so that customers everywhere can enjoy the benefits of these solutions."

About Lionbridge: Lionbridge Technologies, Inc. provides solutions for worldwide deployment of technology

LanguageForce Now Glossa?

LanguageForce, the sometimes controversial vendor of the Universal Translator MT system, appears to have gone out of business and resurfaced again with a new name. A notice appeared this summer on the LanguageForce Website indicating that it had permanently closed its doors in May 2001 and sought relief from creditors. The closure included GoToWorld.

In September, a solicitation for business partners and investors was circulated with a business plan for a new company called "Glossa", whose product is the same "Universal Translator". The company does not appear to have a website yet, and does not appear to be related to a Greek translation company also called Glossa. This is at least the second time that the company has gone out of business and reappeared under another name. The previous change was from WebForce to LanguageForce.

The Universal Translator is notable for its ambitious interlingual design, independently confirmed in a report by Professor Bonnie Dorr of the University of Maryland. The Universal Translator system offers a very large number of language pairs, including many exotic languages. ❑

SYSTRAN Acquires Aurora

[Adapted from press release]

Aurora, a French publisher and integrator of Open Source and Linux solutions, was acquired by the SYSTRAN Group in September. Aurora will continue with its current engineering and services activities and will provide integration for and technical support services to SYSTRAN's worldwide customers, expanding its expertise to multilingual content management for Internet and intranet applications.

"Aurora is a highly recognized integrator and IT services provider in the Linux and Open Source arena in France. Aurora will allow us to include a full range of turnkey, multilingual services in our professional solutions offering to resellers, integrators and partner software publishers," commented Dimitris Sabatakakis, Chairman and CEO of SYSTRAN.

About Aurora

Aurora implements Linux and Open Source technologies to provide businesses with global solutions for multilingual content management and knowledge management projects. Headquartered in Boulogne-Billancourt (92) France, Aurora has recorded 0.8 M Euro as sales in 2000 and 0.7 M Euro during the first half of 2001.

About SYSTRAN

SYSTRAN has been a provider of MT technologies and applications for more than 30 years. In 2000, SYSTRAN's sales amounted to Euro 9.3 million, for an operating margin of 12.2% and a net profit of Euro 0.94 million. SYSTRAN is headquartered in the Paris area (France) with subsidiaries in California and Luxembourg.

For more information, contact SYSTRAN: Frederic Burband, Tel: +33 (0) 1 39 34 97 99, Email: burband@systran.fr, or see: www.systransoft.com. Aurora: Jean-Noel de Galzain, Tel: 33 (0) 1 58 17 03 20, Email: jnoel@aurora-linux.com, or see: www.aurora-linux.com ❑

US Globalization Survey

LISA and Geomarkets announced the results of First Annual US Technology Globalization Survey 2001 on November 1. The study, "2001 US Technology Globalization Survey: Globalization & Localization Services, Markets and Trends" confirms that the localization market is healthy and offers tremendous market opportunities for solutions providers, which will continue to grow within the next three years.

By interviewing over 200 key executives at high technology companies regarding their globalization plans, budgets, preferences, and forecasts, the 2001 US Globalization Technology Survey is a comprehensive, in depth, end-user survey regarding the current and future state of the globalization market, the level of spending allocated, outsourcing requirements, language and market forecasts, and executive opinions on globalization issues. The study was sponsored by Idiom Technologies, LionBridge Technologies, SDL International, Berlitz GlobalNET and ALIS Technologies.

Key Findings

• 66% of US-based IT companies' Websites are English-only while more than half of all Internet users are non-English speaking

• The estimated size of the US market (2001) for localization is $5.85 billion.

• IT companies, on average, spend between 2-3% of their revenue on localization projects, which range between $500K-$1 million, and up.

• Asia-Pacific is the most attractive international region for the IT industry.

• China and Germany are the two countries where the majority of IT companies have established plans to expand into during the next year.

• German and Japanese are the two most popular languages for IT companies' Web site and Intranet localizations.

• The top three challenges facing companies attempting globalization are: understanding indigenous business practices, local representation, and cultural differences/barriers.

The report is aimed at executives and localization vendors; international vice presidents of sales, marketing, and business development who can use it to benchmark their own globalization/localization initiatives against the US technology industry; and managers who want to understand how the process of globalization and localization can be used to gain a competitive advantage.

The report is available from LISA for US$495 for corporate members, and US $890 for non-members. Order at: www. lisa.org/interact/2001/geomarket.html ❏

SDL Lands Clients with Enterprise Translation Server

SDL International, together with its clients announced two implementations in November for its Enterprise Translation Server (ETS) technology. Contrary to expectations that SDL would focus on MT as a component of its localization services business following its acquisition of the Transcend MT system earlier this year, SDL is actively marketing the MT system itself *(see interview this issue with SDL's Hedley Rees-Evans).*

One of the clients is Belga, Belgium's first National Press Agency where MT will support translations of European Union press information. Belga's EU news will be published across various media channels, including the official website of the EU Presidency. SDL's Enterprise Translation Server has been selected as a "first-pass" automatic translation solution, prior to post-editing by the appropriate EU journalists. As part of the agreement, SDL committed to assist Belga in a six month pilot programme, by fine-tuning the dictionary functionality and analysing other integration points for ETS in Belga's news feed and news processing activity. The

U.S. Social Security Administration to Test WebSphere

The Social Security Administration announced its intention on September 20, to negotiate with IBM under the authority of Federal Acquisition Regulation for proof of concept using IBM's WebSphere Translation Server. The WebSphere Translation Server will soon be installed in the SSA's laboratory environment (Electronic Technology Center) server at SSA Headquarters in Baltimore, Maryland, and the proof of concept will take place on site.

This procurement is part of SSA?s Foreign Language Translation Project, the goal of which is to test the effectiveness and practicality of machine intelligence software that is designed to translate English to Spanish and Spanish to English. The Government is currently in the process of acquiring the IBM WebSphere Translation Server software. However, ancillary technical services are also needed in order to stage realistic tests of the full operational capability of the software. The proof of concept will involve examining the software for its suitability to several agency business applications, including but not limited to the following: a) document translation with a view to processes in which 'gist' rather than exact translation would be acceptable; b) electronic message translation for e-mail; and c) agency website translation. Interested user components will devise client service 'scenarios' to represent their specific requirements and to establish performance criteria for accuracy, speed and ease of use. Success or failing results of these tests will determine whether pilot testing will proceed and whether other foreign language translation by machine intelligence will be similarly tried.

If you have any questions regarding this announcement, please contact Kathy Fain, contract specialist, at 410-965-4853. ❏

Sakhr Brings Arabic Online with Two New Services

In October, Sakhr Software, the Egyptian developer of Arabic machine translation systems, announced two new online services for English<->Arabic translation. The first is Ajeeb *www.ajeeb.com* a free Arabic and English translation portal enabling realtime translation of Arabic and English news sources in either direction. The portal service is described in an interview with Fahad Al Sharekh, Chief Executive of the Ajeeb division of Sakhr Software. The interview appeared on November 12 and is available at the Wired News site (*www.wired.com/news/culture* search for Sakhr). Try the service at *tarjim.ajeeb.com/ajeeb/*.

The second is ArabSite, Sakhr Software's Web Arabization solution, which enables users localize any English website without creating an Arabic site.

Customers who maintain English-only websites face growing demand for an Arabic version of their websites to satisfy Arabic-speaking visitors. Simply creating a bilingual English/Arabic website does not solve the problem. The Arabic site must be manually updated every time the original English site changes. Since most sites change frequently, keeping track of all these changes requires a dedicated team, and new hardware investments.

Sakhr's solution provides dynamic constant updating of the Arabic site through a combination of English to Arabic machine translation and human translators. Sakhr's machine translation engine is integrated with the ArabSite local server to optimize the translation process in a controlled work cycle. Human translators play an important role in the review and approval process.

Achraf Chalaby, Sakhr's director of Research, notes "ArabSite generates on-the-fly an Arabic web site for an already existing English site. The English site owner updates his English site and immediately ArabSite senses the changes and updates the Translation Memories automatically. No client intervention is

Products And Applications

Oki Electric Pilots Collaborative MT User Community

Oki Electric Industry Co., Ltd. announced that it has developed a web-based machine translation system that can be improved in terms of accuracy and scope through online collaboration by users. A presentation about the system was made at MT Summit VIII, and appears in the proceedings. A tour of the functionality is available at *www.yakushite.net.*. The site will be evaluated in a trial with the participation of academics and other selected users.

Oki's new system leverages the cooperative efforts of online users for the creation of highly accurate dictionaries, enabling people with deep knowledge of a particular subject to enhance specialized dictionaries for online machine translation. "Until now, machine translation has not been used widely because of its low quality. Oki's Yakushite Net, however, will enable internet users to easily improve the quality of machine translations through collaboration." said Toshiki Murata, team leader of Oki's Service Media Laboratory. Users with a common interest in a specific field can create a "community," with a unique domain on the machine translation website. In addition to the system's basic translation dictionary, the community creates and updates a specialized dictionary with words and phrases encountered in its specific field. The community also manages machine translation results, which includes correcting and modifying translations produced by the system. The system learns new translation patterns based on how users correct and modify the machine's original translation.

MT in Multinational Customer Support

eCRM (Web-based Customer Relationship Management) is a large and growing category of enterprise business support software. Often known in terms of large-scale back-office database applications, with major players like Siebel, this type of enterprise software typically kept track of customer orders, buying behavior, technical support use and other customer-related record keeping. But in the face of burgeoning e-commerce, the name CRM has come to be used for an increasingly wide variety of customer-directed functions, including email interaction with customers (often automated) and self-service technical support management.

Since the announcement of AutoDesk's implementation of SYSTRAN for self-serve technical support, a number of CRM vendors have issued press releases emphasizing that their products are multilingual. In some cases this means something akin to "internationalized" or Unicode compliant, and implies only that the product can potentially work with any language and character set. But one developer, a software company headquartered in Huntington Beach, California has gone so far as to bundle "Language Neutral" MT into their product ParaCRM™. In addition to third party machine translation, a proprietary AI engine wrapped around the MT process improves the quality of translation output by utilizing context beyond the sentence level. The ParaCRM system, now in Version 2.0, was first released in early 2001, and has attracted customers such as Volvo. Two capabilities of the system have received the most enthusiastic response from clients: the MT integration into real-time chat which handles conversational language and colloquialisms well; and MT integration into self-service customer support via free-form nonstandard questions (NSQ) and answers. In fact, CEO Dr. Sergey Aityan has been using the chat capability himself to communicate with a prospective client in Japan.

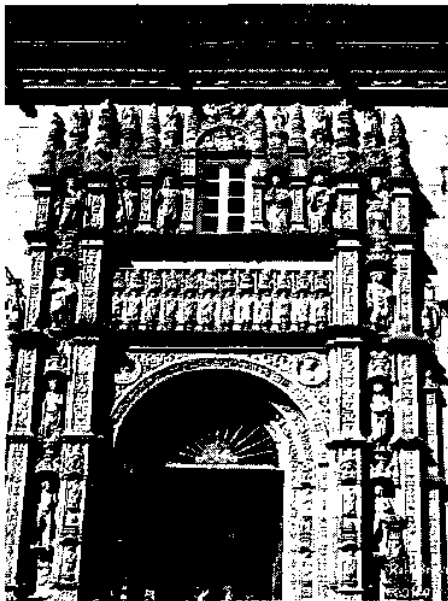Based on the positive reception of the

Conference Reports

MT Summit VIII

Junichi. Tsujii

MT Summit VIII at Santiago de Compostela was the most memorable Summit so far for three reasons, the venue, the program, and the incident in the U.S. on September 11 that immediately preceded the conference.

Venue   *Photos by Ralf Brown*

Hostal de los Reyes Catolicos, where the Summit took place, was magnificent beyond anyone's expectation. The contrast between the ambience of the venue and the city of Santiago could not be starker, when compared with KRDL in Singapore where the last Summit took place. While Singapore showed us a lively, super modern, cosmopolitan and at the same time very Asian bustling ambience, Santiago de Compostela amazed us by its European cultural heritage and its tranquility.

The hotel is a historical building founded in the 15th century to minister to pilgrims. While the public can only see its splendid Plasteresue door, the old chapel and two Renaissance courts flanking the vestibule, the sessions of the Summit took place inside the build-ing and the participants who stayed at the hotel could enjoy every bit of the great European heritage.

While I have trust in language, there are rare occasions when language is not expressive enough. I can only say to those who could not come this time that you should visit the city and stay at the hotel if you ever have the chance.

Program

The second reason is the diversity and richness the conference program delivered. As in the previous Summits, we saw presentations by policy makers, MT vendors, researchers and users of MT systems. The Summit was accompanied by five workshops that addressed diverse aspects of our field, from education on MT to evaluation to a future road map of the field.

Thanks to the program chair, Professor Bente Maegaard, participants obtained a comprehensive and balanced perspective of the field, which is developing rapidly. As a researcher told me, he learned at the conference what would really be required by users and the MT industry. The Summit did not fail to deliver what the previous Summits had delivered, i.e. the mutual fertilization among people who are involved in this field from very different perspectives. Papers by users were not many, but we saw quite a few users at the conference, and since they were vocal, their presence really animated discussions.

From the technical view point, those that had once been called new paradigms of MT, such as Example-based MT, Statistical MT, etc. showed maturity, but still have not reached the stage of replacing the traditional framework in the actual market. The structural complexity of language seems still to prevent these frameworks from delivering working systems in a production environment. However, the maturity will enable these frameworks to be combined with the traditional ones to deliver future MT systems. From the industrial side, we witnessed that predictions made at the last MT Summit were being fulfilled, the predictions that MT would become an integral part of document/information management systems and would be used as a component in a larger system.

World Events

The third reason that made the Summit memorable was the terrible incidents in N.Y. and Washington DC on September 11, just a week before the Summit. Quite a few people, in particular from the U.S., could not help canceling their attendance due to the difficulties in traveling. Not only the people who could not come, but also participants were significantly affected by the incidents. Most of us watched CNN in their rooms. To see the World Trade Center buildings, the symbols of the modern society, collapsing in rooms surrounded by European Renaissance heritages gave us surreal feelings as well as great sadness.

In the opening ceremony, the IAMT award was presented to Dr. John Hunchins who has greatly contributed to our field by his great books and excellent presidency of IAMT, and to whom the great success of this Summit was attributed.   ❑

Summit Workshops

Example-Based MT

Andy Way

This workshop, held September 18 in Santiago de Compostela before the Summit, was organised by Michael Carl, IAI, Saarbruecken, and Andy Way, Computer Applications, Dublin City University. There were 75 attendees, which by far exceeded the expected number of participants. This shows how important the current focus on corpus-based translation research is perceived to be. If further evidence of this were required, in addition to this workshop, there were workshops at EACL/ACL Toulouse, France on Data-Driven MT in July 2001 (see report this issue) and at TALN-01, Tours, France, also in July, on Large-scale MT Applications.

Eight papers were presented at the EBMT workshop in Santiago. These can be classified as follows:

- two 'philosophical' papers, where the characteristics of EBMT systems were debated;
- six 'technical' papers, where particular systems were described.

Philosophical Papers

After the opening remarks made by the workshop co-chairs, the first of the two philosophical papers was given by Davide Turcato & Fred Popowich (gavagai Technology Incorporated, Canada). Turcato & Popowich argued that 'linguistically principled' approaches to EBMT significantly overlap with other 'linguistically principled' non-EBMT approaches. That is, two approaches can be regarded as one and the same approach if they use the same knowledge in the same way. More specifically, any approach to MT will require two sets of information: basic translation equivalences at the word to word level, as well as a repository of phrases which do not translate compositionally. If these knowledge bases are complete for the task in hand, then the linguistic information used by EBMT is indistinguishable

from that of a lexicalist MT system, Turcato & Popowich claim. Harold Somers (CCL, UMIST, Manchester) then presented the second philosophical paper. He examined EBMT from the perspective of Case-based Reasoning (CBR), the idea being to see whether this firmly established AI technique could provide the relatively newer EBMT paradigm with new insights into the solution of the problems of translation. In the main, Somers is pessimistic that there is much to be learned from the CBR community, but adds that perhaps researchers in CBR might find EBMT an interesting special case, given its relatively superficial analysis and representation of the prior translation examples.

Technical Papers

The system descriptions were then presented. Again, these could be classified on two scales: from knowledge rich to knowledge poor on the one hand, and whether they focused more on the working phase of the system or the learning phase. Andy Way (Dublin City University) presented his LFG-DOT models of translation, which combine the linguistic knowledge from the syntactic theory Lexical-Functional Grammar with the probabilistic language models of Data-Oriented Parsing. He showed that unlike other EBMT approaches, LFG-DOT does not suffer from the well known problem of Boundary Friction. Reinhard Schaeler (University of Limerick) then outlined the ideas behind his proposal for a 'Phrasal Lexicon', suggesting (like Turcato & Popowich) that this might be a bridge between Translation Memories, where the user produces the final translation, and EBMT systems which are fully automatic. Kevin McTait (CCL, UMIST, Manchester) presented an approach to EBMT which operates by extracting translation patterns from an aligned bilingual corpus using the relatively 'low level' technique of string matching. The patterns extracted resemble transfer rules to a certain extent, but with fewer constraints.

Ralf Brown (LTI, Carnegie Mellon University, Pittsburgh) combined two techniques from previous work, namely the induction of grammars from (both monolingual and bilingual) unlabelled

MT Evaluation

Building a Community in MT Evaluation: a Report from the Fourth ISLE Workshop at MT Summit VIII

Andrei Popescu-Belis

Time and again complaints about the difficulty of MT evaluation, about non reusable data, and non comparable results resurface at machine translation conferences. At the panel on evaluation hosted by the last MT Summit, one of the most acclaimed proposals for "the most important question an MT evaluator should bear in mind" was simply: "don't I have something better to do with my life?" Nevertheless a more constructive interrogation was equally suggested: "is someone going to listen to what I say?"

Indeed, to avoid not being listened to, is it possible to establish a common ground for MT evaluation (MTE)? The post-summit Workshop on MTE was part of a series of workshops whose organizers believe that users are going to listen to what evaluators say, provided that (a) definitions for the main MT software qualities are agreed upon, and (b) an objective appraisal is given for each metric, stating its domain of use, reliability, statistical behavior, cost, etc.

The MTE workshop at MT Summit VIII was the fourth workshop organized within the framework of the ISLE project (International Standards for Language Engineering), supported by the National Science Foundation in the USA, the European Community and the Swiss Government. The MTE Working Group (which is only one of the ISLE working groups) aims at synthesizing an overview of MTE, which in its final state will comprise a taxonomy of the various qualities and features used in MTE, together with analyses of the metrics that have been used until now. More about this group, mainly based at USC/ISI (Los Angeles, CA) and at the University of Geneva (Switzerland) can be found on its website: *www.issco.unige.ch/projects/isle/ewg*. This site also provides information about the three previous workshops, summarized in an ear-

MT Evaluation Workshop

...continued from previous page

lier issue of the MTNI newsletter by one of the organizers, Flo Reeder (issue 27, vol. 9:3, pp.5-6).

The Santiago workshop took place at the end of a lively conference, in which, despite the September 11 tragedy, numerous researchers from the United States maintained their participation. The summit itself hosted two sessions and a panel on MTE. in addition to the MTE workshop. Together with the high number of participants at the workshop (around sixty), these reflect the vivid interest in evaluation in the MT community. The audience of the workshop was well balanced between users, developers and researchers: the workshop brought together participants from companies. universities. government agencies and international organizations. Inclusion of the workshop proceedings in the MT Summit CD-ROM also promoted its visibility.

As with the previous workshops, this one contained both technical presentations (reviewed papers) and hands-on evaluation applications. Each of the four workshops has had, however, its distinctive focus. In the present case, the need for syntheses of previous experiments and time constraints favored oral presentations and discussions over hands-on exercises. The twelve presentations were grouped into four sessions, and were given at the beginning of each session, the remaining time being left for joint discussion of the session's papers. The themes of the four sessions were: (1) the ISLE taxonomy for MTE and its use; (2) correlations between evaluation measures; (3) analytic measures of output quality, focusing on noun phrases; (4) MTE in relation to other domains.

Hands-on exercises were not altogether absent, since one of the organizers, Eduard Hovy (USC/ISI) proposed to the participants a scoring exercise aimed at estimating the variations of the "fidelity" and "fluency" metrics, applied here by the participants to two texts. The results of this experiment will be described at further workshops, just as results from previous experiments were reported at the present workshop (for

instance in sessions 2 and 3).

The broader goal of this series is the dissemination and discussion of the standardization efforts promoted by the organizers. In this respect, the workshop already presented some interesting uses of the prototype version of the standards, and also paved the way towards more refined analyses of the metrics. Though the main focus is not the proposal of new metrics, the workshop also contained some proposals for a finer-grained distinction of MT software attributes to be evaluated. On the whole, it encouraged the adoption of a common vocabulary and framework for MTE, based on the organizers' proposal, now visible at: *www.issco.unige.ch/projects/ isle/taxonomy2*. The feedback obtained until now, as well as feedback from the community received at the above URL, is currently being processed, before a new series of workshops is initiated.

The organizers of this workshop were: Eduard Hovy (USC/ISI, USA), Bente Maegaard (CST, Denmark), Maghi King, Sandra Manzi, and Andrei Popescu-Belis (ISSCO, University of Geneva, Switzerland) and Flo Reeder (MITRE Corp., USA). The author thanks them for helpful remarks on this summary.

Andrei Popescu-Belis andrei.popescu-belis@issco.unige.ch ❏

Teaching MT Workshop

Francis Bond and Harry Somers

This was the first workshop on teaching machine translation, organised by Mikel Forcada (Universitat d'Alacant) and Derek Lewis (University of Exeter), and held September 22 in Santiago de Compostela immediately following the Summit. There were nine presentations, followed by a panel discussion. Overall, there were two ideas common to most of the papers. The first was that there are two different groups who want to learn about MT: translators and computational linguists. The second was that the web, in particular the availability of on-line MT systems has made teaching MT much easier.

Lorna Balkan (University of Essex) gave an overview of information about MT available on the web. There was an enthusiastic discussion afterwards about whether it was worth trying to set up a site listing resources. The consensus was that it would be better to enhance existing sites, such as the IAMT site, rather than creating new centers.

Dorothy Kenny and Andy Way (Dublin City University) highlighted the differences between the interests of translators and linguists. Translators are *users* of MT software, including translation memories. They are interested in current strengths and weaknesses, as well as hands on training with tools. Computational linguists and computer scientists are *developers* and need to be taught more about the internals of MT systems. Viviane Clavier and Céline Poudat (Université d'Orléans) and Judith Belam (University of Exeter) also emphasized the need for vocational training in MT tools for potential users, contrasted with a more theoretical approach for computer scientists.

Harold Somers (UMIST) added a new group of students with an interest in MT tools: Computer Aided Language Learners. It appears that some students are so convinced of the quality of MT systems that they use them to do their homework! It was suggested that a more proper use would be translating from the second language into the students first, so that they can see what kinds of errors are produced due to structural differences between languages.

Federico Gaspari (University of Bologna) gave the results of an interesting study of translators attitudes to MT before and after a course. In general, after the course, students showed a more balanced and positive attitude to MT. Elia Yuste-Rodrigo (Universität Zürich) also emphasized the importance of considering translators attitudes when teaching them about MT, and noted that the more they knew, the more their attitude improved.

All of the courses on teaching MT took advantage of on-line translation tools. The translator-oriented classes also introduced the students to commercially available translators' workbenches. Juan Antonio Pérez-Ortiz and Mikel L. Forcada (Universitat d'Alacant) gave an example of using multiple on-line systems to show the different translation strategies used.

Data-Driven MT Workshop Assembles Researchers from Different Venues

Jessie Pinkham

The workshop took place in Toulouse, France as part of the ACL/EACL conference on June 7, 2001. It was one of the most highly attended workshops of the conference, with registered attendees in excess of forty. The workshop lasted one very full day, from 8:30 to 5:50 pm.

The workshop was organized by Franz Josef Och (RWTH-Aachen), Kevin Knight (USC/ISI) and myself. With increased availability of online corpora, data-driven approaches have become central to Natural Language research. A variety of data-driven approaches have been used to build Machine Translation systems – example-based (EBMT), statistical MT and other machine learning approaches – and there are all sorts of possibilities for hybrid systems. The workshop focused on Data-Driven MT, by which we mean all approaches which develop algorithms and programs to exploit data in the development of MT, primarily with the use of large bilingual corpora created by human translators, and serving as a source for training data for MT systems.

The purpose of the workshop was to encourage interaction between people working on different flavors of Data-Driven MT. The papers represented the multiple approaches well, but we found that interaction was limited at the workshop due to a very tight schedule and the lack of time for a panel discussion. Some discussion in the question periods made it clear that different approaches should ideally be working from the same data sources, so that the strengths and weaknesses of each could be better compared.

The invited speaker was Herman Ney of RWTH-Aachen. His presentation, entitled "Stochastic Modelling: From Pattern Classification to Language Translation", gave an overview of the stochastic modelling approach to machine translation. He discussed the properties of the components of his system,

ATA Postediting Workshop

This year's American Translators Association conference included a pre-conference workshop jointly organized by the ATA and AMTA in an effort to encourage more sharing and communication between the associations and their members. The workshop was held the morning of Wednesday, October 31 preceding the start of the ATA conference at the Biltmore Hotel in downtown Los Angeles, led by Walter Hartmann.

Walter Hartmann

The focus of the workshop was slightly changed from its original intent of an open panel discussion and shaped into a quick course on MT, its potential for translators, the selection of MT and post-editing. 18 participants signed up for the Workshop to explore the question, "How Can MT Possibly Help a Professional Translator?" Not all participants were translators, there were also members of academia, employees of translation companies looking into the possible use of MT and even an employee of an MT company.

Alan Melby led off with a short introduction of the topic and MT, and he had all participants introduce themselves and explain their interest in MT. In general, the consensus seemed to be that they wanted explore MT in general and the possible use for their own purpose.

Then the presentation began with an overview of the various uses MT has successfully claimed for itself and the proliferation of MT over the last few years in government, institutions and corporations. The increasing role of MT on the Internet that gave it renewed vigor over the last years was considered as well. What this means (or could mean) to the translator, be it freelance or in-house, was discussed next.

The approach to MT will vary depending on whether one works within a corporation with MT systems, or as a contractor to translation companies, which are using MT in increasing numbers, or

HT Evaluation at ATA

You can't manage what you can't measure

This year's American Translators Association conference, held in Los Angeles, October 31 through November 3, included two presentations on methods and initiatives for evaluation of translations.

The first, known as SAE J2450 (organized through the Society of Automotive Engineers), is a methodology for assigning a numeric quality rating to a translation. It defines a framework for classifying errors (wrong term, syntactic error, etc.), which can be ranked for severity. The metric is much like those used internally by MT developers to target system strengths and weaknesses. Fred Meinberg, of Techworld Language Services presented the standard on behalf of Rick Woyde, President of Detroit Translation Bureau, and chairman of the task force that developed the metric.

A significant part of the value of such a standard is that it provides the means to objectively discuss price, turnaround and quality with translation consumers. For example, when a job is being sold, agency and client can agree on a numeric quality rating for the finished product which characterizes the approximate number and severity of errors that may be found in the target text. The client has a clear means of deciding what level of quality they are willing to pay for. The translation provider has a clear, measurable target.

The beauty of the metric in practice is that it only requires detailed evaluation of small samples, which makes it affordable to use on every job. Evaluation samples were as small as 300 words drawn from several locations in a several-hundred-page auto maintenance manual. See the website for more information: www.sae.org/technicalcommittees/j2450p1.htm

The second evaluation method was presented by representatives of two localization firms, Franco Zearo of Lionbridge, and Riccardo Schiaffino of J.D. Edwards. The presenters did not go into detail on what should be measured or

Special Feature: Speaking of MT

A Chat with SDL International

In the last year three of the most established and respected U.S. MT systems have changed hands. Among them was Transcend, which sold to SDL International, an international localization company, headquartered in the UK (reported in #27). SDL is among the worlds largest localization companies, with approximately US$50 million in annual revenues. SDL is publicly traded on the London Stock Exchange.

The Transcend MT system has had as varied a history as any MT system in existence. It started out as Weidner MicroCAT in the early 1980s, and was later bought by Intergraph (a U.S. CAD company). Transparent Language, a US developer of computer-aided language learning software, bought Transcend in 1996, and then sold it in February of this year. The "Transcend" name is retained on just the desktop product at SDL. Personal and enterprise products have taken on new designations, as reported below. While plans for ongoing development and deployment of the other recently sold systems (Logos and the Mendez iTranslator) have not been announced, SDL has regularly made announcements on their integration of MT into localization tools. We wanted to find out more about the use of MT at SDL, and plans for the future. MTNI spoke with Hedley Rees-Evans, Marketing Director, and Keith Mills, CTO at SDL.

MTNI: For many years the localization community rejected machine translation as inadequate to the task of localization. Is the localization world ready for machine translation now?

Rees-Evans: We debated this before the acquisition. In the past, there was a very similar resistance to the adoption of translation memory. Translators said that their expertise was contained in their heads and that couldn't be replicated through software. However, we believe that customers deserve the most cost-effective solution to their needs – if

that involves the use of MT or TM, then so be it.

We make MT available in various ways. One is as "Enterprise Translation Server" – 'ETS'. This can be used in the context of something like Lotus Sametime or Microsoft Office. This is an application area where localization is just inappropriate - where human translation isn't possible, due to the speed of volume of the throughput. SDL has a consultative approach for customers with a requirement for high volume international communications like this, so we provide dictionary development to improve the quality of their translations. We analyze sample data and add the appropriate terms. This is something that differentiates SDL – the breadth of our customization service.

SDL offers several levels of MT products. Easy Translator is inexpensive and doesn't include customization.

Transcend is a desktop product that allows individual users to do customization. ETS is a bigger enterprise level product which comes with support and optional services. ETS competes with Websphere – Translation Manager, SYSTRAN and iTranslator.

MTNI: In August, SDL announced the new product – "Localization Suite" which incorporates the Transcend MT system. Was it in fact released in September as planned? How has the MT component been received by the localization community?

Rees-Evans: The first step was to integrate ETS into SDLWebFlow. (SDLWebflow is a multilingual content management system which significantly automates project and file management for web localization projects, as well as integrating Translation Memory and quality review.) We put ETS alongside translation memory, so that when a translator is presented with a segment or sentence that is missing from the Translation Memory, or with an inadequate match, they are offered MT output.

Mills: The SDL Localization Suite, which includes all of this functionality has been available for 3 months. The localization tools outside of MT have been available for several years.

Rees-Evans: Some documents require high quality output, and it's hard to see how MT can play a role there. Most customers require a particular style. The benefit is in getting the MT and Translation Memory to work together with user dictionaries. If the dictionaries reflect the user's style, you get the consistency desired. Then you can accelerate the localization process and give value. Our R&D department is working on this direction.

Translators vary (in their reaction to the availability of MT in SDL products). SDL has lots of technical project managers and engineers who get exposure to the products first, and their feedback is included in the development process. Clients are interested in anything that will reduce their time to market.

MTNI: In observing the localization market, SDL seems to be one of the more technically savvy companies. (SDL develops and markets the widely-used translation memory system SDLX in addition to SDLWebFlow, mentioned above.)

Rees-Evans: We have made very practical investments in technology. This is a long term strategy that has included investments and acquisitions.

MTNI: How big is SDL?

Rees-Evans: SDL (Which is headquartered in Berkshire, U.K.) has a total of about 700 people worldwide, working in 25 offices. Weekly translation output is about 3 million words per week. About 7% of our people are involved in technical development, including computational linguists developing MT. Annual revenues are about US$50 Million per year (SDL is a publicly quoted company on the London Stock Exchange) ❑

A Chat with Young Researchers

We wanted to get a fresh perspective on the future of MT, so we talked to some of the people who will be creating that future: two PhD students in computational

linguistics at Carnegie Mellon University: Kathrin Probst, and Ying Zhang, who goes by the nickname "Joy".

MTNI: Is there a particular approach to MT that you think is most promising or that inspires you?

Kathrin: I believe that the future of MT lies in rapid-deployment methods. In the past, researchers have spent years and years writing rules for machine translation systems. If MT wants to branch out to new languages, it would be very useful to have tools that allow for MT systems to be built fast. So, for instance, I believe that statistical and example-based methods are going to be used more frequently. Also, I hope that methods that apply machine learning techniques to learn rules are going to be explored more.

Joy: I view the building of a machine translation system as an automated learning process. Take Google as a related example. Google ranks documents highly that are not only relevant to the query, but are also referred to by many other sites. In this way, human knowledge or judgement is automatically extracted. In machine translation, data-driven approaches basically strive for something similar. However, I am not sure that we can obtain enough knowledge from a bilingual corpus. In other words, pure statistics cannot solve all problems.

Kathrin: I agree that statistics cannot extract all the knowledge an MT system needs. I think that in theory the most promising systems are those that combine data-driven approaches with linguistic knowledge. However, we do not yet know very well how to build such systems successfully.

Joy: I think that natural language can be seen as a social phenomenon. For instance, even if you know grammar of a language and have good vocabulary, if you don't live in an environment or community in which the language is spoken, you will form sentences that native speaker would not form. Thus, the development of natural language is a collaborative effort of people. If we can develop a system that can interact with people and learn from people by

updating its knowledge, it would be very promising one. However, this is hard. For instance, we will have to assume that the computer has a certain degree of world knowledge.

MTNI: Do you think that there are particular applications of MT that will bring MT broader acceptance and/or give people the confidence to user it for even important/sensitive communications? What will make this possible?

Kathrin: I think that in the near future, there will be no general-purpose MT systems that have a high enough quality to be used as stand-alone systems for sensitive or important communication. On the other hand, I think that MT will be used more and more as an underlying technology for other applications such as Information Retrieval and Extraction, Question-Answering, etc. The economic growth of these applications will also lead to an economic growth of MT.

Joy: I agree with Kathrin. I don't think that in the near future we can provide MT products that people will feel comfortable using in important communications. I think MT can be used in the future in cases where people have additional information sources, e.g. in caption translation where you have visual information, in sign translation where you have the additional visual clues about the environment the sign is in, etc.

MTNI: What part of either the research process or your visions for the future help you to get through the often slow and discouraging process of developing and refining theory?

Joy: I see MT as a personal challenge. When I am confronted with a problem and the solution is obvious to me from the start, the problem is not interesting. It's like with rock climbing, where you want to climb the most challenging route. When you manage to do it, you feel personally satisfied.

Kathrin: Not having been in research for long, I still have visions and dreams. My dream is to develop a theory or system that will have impact on the field of MT in general. A system with which we can build MT systems for any language pair, fast, and with little data.

At the moment, I am working on a system that makes use of machine learning techniques to (semi-) automatically infer syntactic transfer rules. If such a system can be used for any language pair, MT could branch out to language pairs that have not been studied, such as minority languages.

MTNI: What have been some landmark events in your personal history with the field?

Kathrin: When I first went to college, I started studying linguistics. The field was very interesting to me, but I did not want to stay in theory only. So I started a major in computer science, with the plan to combine the two fields in graduate school. I was especially interested in multi-lingual applications because I have studied languages all my life. In my small liberal arts college, I was the only person who was interested in this research field. When I started my studies at Carnegie Mellon University, I was surprised and pleased to see that there are lots of other people who share my interest!

More recently, the MT Summit this past September was a landmark event. Learning what many other people in the field work on and think about very much motivated me.

Joy: This April or March, I visited an exhibition in Lexington, MA. It's called "The History of the Bicycle". I was surprised at how people came up with the idea of using two wheels for a vehicle instead of three and four, which is more intuitive. Also, during the visit I was reminded of how people invented airplanes. First, people tried to mimic birds with moving wings. That didn't work. Nowadays, all the airplanes have fixed wings, unlike birds. Eventually, people managed to build something that was very useful although it did not work like the model did. I am hoping that the same will happen for MT.

Also, some successful products turn out to be not used for what they were initially designed for, e.g. Interstate highways, which were initially designed by the army to move troops faster. Maybe the future will bring an application of MT that we have not even thought about. ❏

Association News

IAMT and the Regional Divisions: EAMT

This column is the second in a 4-part series on the IAMT and its regional divisions. We look at each association's activities and organization, and do a profile on an influential member of that organization. This time, we look at the EAMT (European Association for Machine Translation). The EAMT was founded in 1991, at the same time as AMTA, AAMT (or JAMT as it was originally called), and IAMT. The first president, and author of the constitution of EAMT, was Maghi King. The association is based in Switzerland.

EAMT Workshops: Held annually, except in years when the MT Summit is in Europe, the workshop is typically a two-day event, consisting of submitted papers and invited presentations on a theme related to MT. The proceedings are published by EAMT. The full proceedings for the first and third workshops (Vienna, 1996 and Geneva 1998) are available on the EAMT website at: www.eamt.org/archive.html, as is the table of contents for the fifth workshop (Ljubljana, 2000). The next one is tentatively scheduled for May, 2002 to be held before CLAW (the Controlled Language Applications Workshop). Check the EAMT website for details.

General membership meeting: A general assembly is held at the annual EAMT Workshop (or the Summit, when in Europe). The meeting covers association business, and includes election of officers. The current slate of officers were elected at the recent MT Summit in Spain.

Articles of Association: Viewable at www.eamt.org/archive/statutes.html

Bibliographic services: John Hutchins offers a bibliographic service to EAMT members. For a nominal fee, members can order photocopies of articles on MT publications in the EAMT archives which are under his care: WJHutchins@compuserve.com

MT-List: The MT-List serves as a forum for the discussion and exchange of information for all aspects of translation technology. The MT-List is moderated by EAMT Webmaster, Colin Brace. Subscription is open to the public . Subscribe or scan the archives at: www.eamt.org/mt-list.html. ☐

EAMT People

John Hutchins, President of the EAMT, and immediate past-president of IAMT, was interviewed in MTNI last year (Issue #24, Spring 2000). This time, we get to know Bente Maegaard, EAMT executive board member, and program chair of MT Summit VIII.

Introducing Bente Maegaard

Bente Maegaard entered the new field of computational linguistics in the early 1970's after completing a MSc in Mathematics and French and the University of Copenhagen, Denmark. When she became aware that a field existed which matched her unusual background, she thought, "That's for me!" In fact, she was one of the pioneers of the field in Denmark.

Eurotra

After teaching in the Department of Applied and Mathematical Linguistics for several years, she and a growing group of researchers in computational linguistics banded together to form the Danish team for EUROTRA in 1984. This was Bente's first exposure to machine translation. Her previous work focused on corpus frequency studies and grammatical analysis of Danish. Bente was elected chair of the EUROTRA Liaison group (the committee consisting of one member from each national team) in 1986 and led the committee till the end of the project in 1992.

The Danish EUROTRA group was quite large and very active. At its peak, the Danish group had 26 or so people out of the 200 participating in EUROTRA in all of Europe. The result of EUROTRA was a prototype MT system treating each of the 9 official languages of the EU at that time (Danish, Dutch, English, French, German, Greek, Italian, Portuguese, Spanish) and a good deal, but not all of, the 72 language pairs.

People are sometimes surprised that little commercial software came out of the EUROTRA project. Maegaard thinks that the problem was a lack of

EAMT Officers	Officers are elected for three year terms during General Assembly meetings

Executive Committee

John Hutchins	Norwich, UK	president	WJHutchins@compuserve.com
Viggo Hansen	Copenhagen, DK	secretary	vha@hofman-bang.dk
Tony Clarke	Basel, CH	treasurer	Anthony.Clarke@cls.ch
Jaro Lajovic	Ljubljana		Jaro.Lajovic@mf.uni-lj.si
Bente Maegaard	Copenhagen, DK		bente@cst.ku.dk
Colin Brace	Amsterdam, NL	webmaster	cb@lim.nl

Co-opted members (appointed by the executive committee for special duties)

Jörg Schütz	Saarbrüken, DE	MTNI Regional Editor	joerg@iai.uni-sb.de
Harold Somers	Manchester, UK	MT Journal Liaison	harold.somers@umist.ac.uk

industrial partners. There was no clear path or leadership for commercializing the results of all the work. After building their research prototypes, the researchers were ready to move on to the next topic, and it is a long way from research prototype to a commercially viable product. In Denmark, however, a patent law firm, Hofman-Bang, heard about the Danish research group and proposed scaling up the English-to-Danish prototype for patent translation. The resulting system, PaTrans (pronounced "pay-trans") has been in active use since 1994 and translates approximately 400,000 words per month. The system was tailor made for Hofman-Bang (now Zacco) and is used in-house by their translation company Lingtech. The system still receives technical support and ongoing development from the staff at the Center for Sprogteknologi (Center for Language Technology), where Ms. Maegaard has been the director since its creation in 1991.

Center for Sprogteknologi

The Center for Sprogteknologi is a government research institute in Copenhagen, Denmark. Although it is housed on the University of Copenhagen campus, the Center is not administratively affiliated with the university, but reports to the Danish Ministry of Information Technology and Research. There is, however, a strong symbiosis between the university and the Center. Researchers from the Center frequently teach courses at the university, and the Center employs students, and takes advantage of the opportunity to interact with university faculty. In addition to PaTrans and some other commercial contracts, the Center has ongoing research activities in lexicon and grammar building, ontology building, concept-based IR using the ontologies, machine translation and integration of language technology into multimedia environments. The center currently consists of 21 people primarily computational linguists and computer scientists. In spite of the recent availability of natural language tools in software and on the Internet, there is still a relatively low level of public awareness and understanding of language technology, and thus another part of the Center's mission is public education and outreach. Periodic public seminars are scheduled, for example Center staff re

cently presented an overview of MT Summit VIII for the general public, which was attended by students, MT users, translators and agencies.

MT Summit VIII

Reflecting on the recent conference in Spain, Bente has many positive recollections, and a well-deserved feeling of satisfaction. We asked her to comment on the experience:

MTNI: What was your general impression of the conference?

Maegaard: The success of a conference depends on a couple of things that one cannot really control. The first is the quality of the submissions one has to work with. We were fortunate that we had a very good selection of submissions to work with in building the program. The second is the general mood or ambience of the conference. I don't know how it gets created. You can't force it to be positive, but my impression was that the ambience of the conference was very good. People seemed to be talking to each other and enjoying themselves. All of the sessions, particularly the panels, had very lively audience participation and discussions following presentations. The fact that we have this tradition of bringing together users, researchers, and commercial developers is wonderful and provides a very fertile environment for exchange of ideas.

MTNI: In reviewing submissions and putting the program together, were there any surprises to you in terms of the international state of MT?

Maegaard: No surprises, but I was very happy to see quite a number of submissions on dictionary development and knowledge acquisition. This is the first time that there has been such a number of papers on terminology and word and phrase extraction from parallel corpora, alignment, and other techniques that

support automated lexicon building. It is not exactly a new topic, but it is a very important one. I think the only way for the field to move forward is through automatic or semi-automatic knowledge acquisition.

MTNI: Every conference seems to present some complications that make the job more of a challenge than anyone expects when they sign on as organizers. Your program co-chair (Jo Lernout) was forced to withdraw early on.

Maegaard: That was a pity. I was very sorry not to have input from the viewpoint of industry. When we signed on as program co-chairs I felt very good about having an industry-academic pair in that role, and the loss of a more commercial perspective was the main disadvantage. But I think that the program was fine anyway, and well balanced.

A rather important challenge was the September 11 attack on the U.S. At the very last minute 30 or 35 registrants dropped out. In addition, 6 or 7 papers, as well as a post-conference workshop were cancelled, which required some on-site adjustments to the schedule. At the same time, three or four participants who were not originally on the program came forward with presentations they happened to have prepared, and they jumped into some of those program slots. So there were some unexpected opportunities.

MTNI: Sometimes there are also serendipitous events that are a boon to the organizer. Were you so lucky as to have any of those?

Maegaard: No single event, but a number of circumstances made the conference very smooth and pleasant. First, the weather in Santiago, which is often rainy at that time of year, was clear and sunny. Second, the conference site (Hotel de los Reyes Catolicos) as well as the historic city of Santiago de Compostela were beautiful and added a special atmosphere. Finally, the organizing committee members worked very effectively together. John Hutchins as the general chair and I were always around and visible to deal with any issues that arose. Jörg Schütz pulled together a very lively group of exhibits, including academic exhibits. And Carlos Paz worked very hard and was very effective as local arrangements chair.   ❏

Conferences and Events

AMTA 2002
Tiburon, California
October 8-12, 2002

Preliminary Call

The Association for Machine Translation in the Americas (AMTA) is pleased to announce its fifth biennial conference, planned for October 8-12, 2002, in Tiburon (near San Francisco), California.

From Research to Real Users

Ever since the showdown between Empiricists and Rationalists a decade ago at TMI-92, MT researchers have hotly pursued promising paradigms for MT, including data-driven approaches (e.g., statistical, example-based) and hybrids that integrate these with more traditional rule-based components.

During the same period, commercial MT systems with standard transfer architectures have evolved along a parallel and almost unrelated track, increasing their coverage (primarily through manual update of their lexicons, we assume) and achieving much broader acceptance and usage, principally through the medium of the Internet. Web page translators have become commonplace; a number of online translation services have appeared, including in their offerings both raw and post-edited MT; and large corporations have been turning increasingly to MT to address the exigencies of global communication. Still, the output of the transfer-based systems employed in this expansion represents but a small drop in the ever-growing translation marketplace bucket.

Now, 10 years later, we wonder if this mounting variety of MT users is any better off, and if the promise of the research technologies is being realized to any measurable degree. In this regard, we pose the following questions:

Why aren't any current commercially available MT systems primarily data-driven?

Do any commercially available systems integrate (or plan to integrate) data-driven components?

Do data-driven systems have significant performance or quality issues?

Can such systems really provide better quality to users, or is their main advantage one of fast, facilitated customization?

If any new MT technology could provide such benefits (somewhat higher quality, or facilitated customization), would that be the key to more widespread use of MT, or are there yet other more relevant unresolved issues, such as system integration?

If better quality, customization, or system integration aren't the answer, then what is it that users really need from MT in order for it to be more useful to them?

Call for Papers

We solicit participation on these and other topics related to the research, development, and use of MT in the form of original papers, demonstrations, workshops, tutorials, and panels. We invite all who are interested in MT to participate, including developers, researchers, end users, professional translators, managers, and marketing experts. We especially invite users to share their experiences, developers to describe their novel systems, managers and marketers to talk about what is happening in the marketplace, researchers to detail new capabilities or methods, and visionaries to describe the future as they see it. We also welcome and encourage participation by members of AMTA's sister organizations, AAMT in Asia and EAMT in Europe.

Details regarding the conference, including submission guidelines, will be provided shortly on the AMTA Web site: www.amtaweb.org

Elliott Macklovitch, General Chair
Stephen D. Richardson, Program Chair ❑

AMTA 2002 Important Dates	
Submission deadline	April 15, 2002
Notification to authors	May 31, 2002
Final versions due	July 15, 2002

HLT 2002
March 24-27
San Diego, California

Call for Papers

Human language technology (HLT) incorporates a broad spectrum of disciplines working towards two closely related goals: to enable computers to interact with humans using natural language, and to serve as useful adjuncts to humans in language understanding by providing services such as automatic translation, information retrieval and information extraction. Following the great success of HLT 2001, the Conference and Program Chairs invite submissions for HLT 2002 from researchers in computer science, linguistics, engineering, psychology, etc., who are exploring innovative methods for improving human language technology.

The Conference will include peer-reviewed research presentations, posters, demonstrations, panel sessions and time for discussion. We expect it to include a session of invited "best of" papers from conferences of sponsoring organizations focusing in particular subdisciplines of HLT. HLT 2002 will also include a special focus on Language Processing of Biological Data, which includes both Information Extraction of Biological Data and Language Modeling of Biological Data, an emerging research area involves a linguistic/language processing view of biological data from the perspective of bioinformatics. The purpose of this special focus within the HLT2002 context is to bring to the attention of a wide audience of HLT researchers the research opportunities and recent breakthroughs in these newly emerging areas. HLT submissions outside of the special focus should be in any area of advanced HLT research.

HLT 2002 will be held at the Catamaran Resort in San Diego, California. The Conference is limited to 330 par-

ticipants. Space will automatically be reserved for authors of accepted papers, posters, and demonstrations.

Organizers: Mitch Marcus, University of Pennsylvania (general chair); David Yarowsky, Johns Hopkins University (co-chair); Clifford Weinstein, MIT Lincoln Laboratory, and Bob Younger, SPAWAR Systems Center (demonstration co-chairs); Aravind Joshi, University of Pennsylvania (special focus chair); Lynette Hirschman, MITRE (special focus co-chair).

HLT 2002 is sponsored by several U. S. government agencies, including DARPA, NSF and ARDA. We are currently arranging sponsorship of U.S. and international research organizations in the range of human language technologies for this and the continuing series of HLT conferences.

See hlt2002.org for details ❑

HLT Important Dates	
Extended abstracts due	January 7, 2002
Notification to authors	February 11, 2002
"Notebook papers due"	March 11, 2002
Final papers	April 22, 2002

SNLP-O-COCOSDA
Hua Hin, Prachuapkirikhan, Thailand—May 9-11, 2002

Call for Papers

The Fifth Symposium on Natural Language Processing (SNLP) and Oriental COCOSDA 2002 will be jointly held in a single event called SNLP-O-COCOSDA 2002. This international conference is hosted by Thammasat University in cooperation with Thailand's National Electronics and Computer Technology Center (NECTEC).

SNLP is an international conference held biannually since 1993 with the cooperative effort of a number of universities in Thailand. The purpose is to promote research by bringing together researchers and practitioners in Natural Language Processing and various related fields. Oriental COCOSDA is an international workshop held annually by the oriental chapter of The Interna-

ACL 2002
Philadelphia, Pennsylvania
July 7-12, 2002

Call for Papers and Proposals

The Program Committee of the 40th Annual Meeting of the Association for Computational Linguistics invites proposals for tutorials and submission of technical papers.

Organization: Pierre Isabelle, XRCE Grenoble, France (general chair); Eugene Charniak, Brown University, USA, and Dekang Lin, University of Alberta, Canada (program co-chairs); Martha Palmer, University of Pennsylvania, USA (local arrangements); Hwee Tou Ng, DSO National Laboratories, Singapore (tutorials chair) nhweetou@dso.org.sg; Walter Daelemans, University of Antwerp, Belgium (workshops chair) Belgiumdaelem@uia.ua.ac.be

See www.acl02.org for links to the full call for papers, and the tutorials and workshops pages. ❑

tional Committee for the Coordination and Standardization of Speech Databases and Assessment Techniques for Speech Input/Output. The fifth workshop will be held in Thailand and incorporated into SNLP-2002. Papers are invited on substantial, original, and unpublished research on all aspects of computational linguistics.

Invited Speakers include Professor David Yarowsky, Johns Hopkins University; Dr. Fangxin Chen, IBM China Research Laboratory; Professor Hiroya Fujisaki, Science University of Tokyo; Professor Hozumi Tanaka, Tokyo Institute of Technology.

Related Events: SNLP 2002, Student Session kind.siit.tu.ac.th/snlp-o-cocosda2002/

See: kind.siit.tu.ac.th/snlp-o-cocosda2002/for details. ❑

SNLP Important Dates	
Submission deadline	February 1, 2002
Notification to authors	March 8, 2002
Final papers	March 29, 2002

ACL Important Dates	
Tutorials	
Proposal deadline	January 11, 2002
Notification on tutorials	January 25, 2002
Descriptions due	March 29, 2002
Course materials due	May 10, 2002
Tutorials conducted	July 7, 2002
Papers	
Registration deadline	January 25, 2002
Submission deadline	February 1, 2002
Notification to authors	April 8, 2002
Final papers	May 10, 2002
Main Conference	July 8-10th, 2002
Workshops	

TALN 2002
Nancy, France
June 24-27

Appel à contributions

La 9ème conférence annuelle TALN est organisée sous l'égide de l'ATALA (Association pour le Traitement Automatique des LAngues) et se tiendra conjointement à la conférence pour jeunes chercheurs RECITAL 2002 et aux JEP 2002 (Journées d'Etudes sur la Parole), appels à communications séparés. La conférence TALN 2002 comprendra des communications scientifiques, des conférences invitées, des séances de démonstration et posters, ainsi que des tutoriels qui, pour ces derniers prendront place lors de la dernière journée de la conférence (appel à propositions séparé à venir). Les langues officielles de la conférence sont le français et l'anglais. Les articles soumis seront rédigés en français (ou en anglais).

See www.loria.fr/projets/TALN ❑

TALN Important Dates	
Tutorials	
Tutorial Proposal	January 31, 2002
Notification	February 8, 2002
Course materials due	May 10, 2002
Papers	
Submission deadline	February 15, 2002
Notification to authors	April 19, 2002
Final papers	May 10, 2002

Reminder: 9th TMI

Keihanna, Japan
March 13-17

Organizing committee: Sergei Nirenburg, Computing Research Lab, NMSU, USA (General Chair); Teruko Mitamura and Eric Nyberg, Carnegie Mellon University, USA (Program Co-Chairs); Francis Bond and Hiromi Nakaiwa, NTT Communication Science Laboratories, Kyoto, Japan (Publicity and Local Arrangements).

For more information, see: www.kecl. ntt.co.jp/events/tmi ❏

EBMT Workshop

...continued from page 7

text and the use of clustering techniques in EBMT to convert translation examples into templates. Brown showed in a preliminary experiment that the amount of training text required for translation from French to English can be reduced by a factor of 12. Michael Carl (IAI, Saarbruecken) presented an algorithm for the generation and filtering of structurally analogous translation grammars from treebanks. He showed that the induction of such grammars can lead to disambiguation of meaning as well as correction of bracketing errors. Finally, Stephen Richardson presented joint work with Arul Menezes (both Microsoft Research, Richmond WA) on automatically extracting transfer mappings from bilingual corpora. Their algorithm was shown to significantly improve the quality of the mappings extracted by using a small alignment grammar in addition to the usual alignment techniques. It was interesting to hear what Microsoft are doing in the general area of MT, both in this workshop and in the main conference sessions.

As well as these papers on EBMT, there were other papers in the general programme on EBMT systems, as well as exploiting Translation Memories at a sub-sentential level (cf. Schaeler, above). The workshop ended with a general discussion session on EBMT inspired by many of the ideas heard

HT Evaluation at ATA

...continued from page 9

how, since these points were presumed to be dependent on the type of translation job. However, they expanded on the theme of representative sampling of translations, including a number of methods for validating the results of detailed evaluations on very small samples of a total text.

In addition to evaluation, Zearo and Schiaffino discussed the cost of repairing errors. Translation is a multi-stage process consisting of content creation, translation, editing and proof reading. The quality at each stage affects the quality and effectiveness of the next. At each stage, the text has more value-added, and correcting errors becomes more expensive, therefore, problems should be addressed as early as possible. Ideally this happens in the content creation stage, producing texts that are more likely to yield good translations.

Given the community's abiding interest in MT evaluation, and the sometimes fiery debates that spring from differing motivations in performing MT evaluations, it was very interesting to see how market forces are shaping evaluation in the (predominantly human) translation and localization industries. Both of the methods presented were clearly quite new, and have not achieved the status of common practice, though the SAE metric has already been the topic of two recent automotive industry events in Europe. It will be interesting to see how these evaluation methods fare in the competitive commercial translation world. ❏

---

throughout the day. Given the level of interest that the workshop generated, two publishers expressed enthusiasm for a book on EBMT based around the workshop proceedings—this we hope will be published in 2002. Finally, the workshop organizers expressed their wish that an EBMT-II Workshop might be held some time in the future, possibly at the next MT-Summit in New Orleans in 2003.

Andy Way is a Senior Lecturer at the School of Computer Applications, Dublin City University. Email: away@compapp.dcu.ie ❏

Data-Driven MT Workshop

...continued from page 9

and reported results on the translation of spoken dialogues in the VERBMOBIL project, by which the stochastic model outperformed other systems in an independent evaluation.

There were a large number of other statistical MT papers: Mark-Jan Nederhof (University of Groningen) discussed the approximation of a context-free transduction by means of weighted rational transduction. Franz Josef Och (RWTH Aachen) described an efficient A* search algorithm for machine translation. Sonja Niessen (RWTH Aachen) described improvements to a German-English statistical translation system brought about by usage of hierarchical lexicon models and the introduction of equivalence classes in order to deal with the complexity of German morphology. Federico Prat's work (Universitat Jaume I de Castelló) presented and discussed a system that incorporated an automatic method of creating regular grammars from corpora, called ECGI, and Loco_C, a particular statistical model for dealing with Grammar Association. Ongoing research on automatic extraction of a bilingual English-Japanese lexicon using three different N-gram models was presented by Kaoru Yamamoto (NAIST). Masaki Murata (Communications Research Lab) gave a report on experiments using a variety of machine-learning methods in the context of translation of Tense, Aspect and Modality in Japanese-English.

Example-based MT systems were less well-represented. Eiichiro Sumita (ATR) described a new approach under this paradigm by which the system retrieves the most similar example by carrying out DP-matching of the input and example sentences while measuring the semantic distance of the words. He demonstrated improved coverage over a commercial MT system in the restricted domain of a phrasebook task for foreign tourists.

Benoit Lavoie's presentation (CoGenTex), also in the general EBMT paradigm, detailed an approach to inducing lexico-structural transfer rules from parsed bi-texts using syntactic pattern matching, statistical co-occurrence and error-driven filtering. The system discussed was for Korean-English. This work was similar in spirit to the work of Arul Menezes for the MT system from

Microsoft Research.

Hybrid systems were discussed in several papers by Microsoft Research and the Butler Hill Group presented by Steve Richardson, Arul Menezes, Bob Moore and myself. The MSR-MT system was described as a combination of rule-based analysis and generation and an automatically created database of transfer patterns stored using the Microsoft Mindnet technology. The system also makes use of statistically derived bilingual word lists, using algorithms developed by Bob Moore. In the main overview paper of the Spanish-English system, the authors reported evaluations of the system trained on technical data that are rated better than an untrained commercial benchmark system. Members of the audience pointed out that they would like to see the MSR-MT system compared to a fully-trained commercial system, and the researchers agreed that this was a necessary next step.

The overall tone of the workshop was very technical, and rightly so. It was refreshing that two of the papers were a bit lighter in tone, and more easily accessible to a general audience. In the general domain of statistical MT, Ulrich Germann (ISI) reported on the difficulties of building a statistical MT system from scratch, including the creation of a small parallel Tamil-English corpus. The project did not succeed in creating the bilingual corpus of 100.000 words it hoped for, in part because of some very important issues: Good translators are hard to find. Inexperienced translators may be cheap, but translation is too hard a task for them. Translators will enjoy reading about this ambitious project.

Also more accessible, Masaaki Nagata (NTT Cyber Space Laboratories) reported on a system to extract English translations for a given Japanese term by collecting and scoring translation candidates from the Web. He showed that there are many partially bilingual documents on the web that can be useful for term translation, discovered by using a commercial technical dictionary and an Internet search engine. The use of partially bilingual text, where English translations appears in parenthesis next to the Japanese technical terms, proved to be an innovative step.

The workshop was a success, and we hope to see additional workshops on DDMT in the future.

Jessie Pinkham, Microsoft Research jessiep@microsoft.com ❏

Lionbridge

...continued from page 3

and content to global 2000 companies in the life sciences and financial services industries, including testing and compatibility services, globalization and multilingual content management technologies. Based in Waltham, Mass., Lionbridge maintains facilities in England, Ireland, The Netherlands, France, Germany, China, South Korea, Japan, Taiwan, Brazil and the United States.

See: www.sail-labs.com, and www. lionbridge.com. Or contact Sara Buda Sara--buda@lionbridge.com or Christiane von Stein Christiane. vonStein@sail-labs.de ❏

Teaching MT Workshop

...continued from page 8

Etienne Blanc (GETA) presented an interactive hypertext environment that can be used by students to see the internals of an MT system. This kind of access to internals is particularly important for computationally oriented classes.

The workshop ended with a panel discussion. In this, the participants decided to set themselves up as a special interest group. The group are in the process of creating a website to serve as a locus for resources on teaching machine translation, including a bibliography. This will be linked to from the workshop website: *www.dlsi.ua.es/tmt.*

In my own experience of teaching MT to linguists. I had also found the use of on-line resources invaluable. Students found the process of translating a text, looking at the translation, editing the text slightly and trying again, not only instructive but fun. It would be nice if there were more systems available on-line which also gave some access to the system internal representations. One thing that surprised me was that there was almost no mention of open-source software and dictionaries. For teaching computational courses, actually modifying existing systems seems invaluable, although maybe too much to expect in a short course.

Francis Bond, NTT, bond@cslab.kecl. ntt.co.jp. Harry Somers, UMIST, Harold.Somers@umist.ac.uk ❏

ATA Postediting Workshop

...continued from page 9

even if one uses MT oneself for freelance work. There are advantages, disadvantages and even pitfalls in all three cases. In general, however, given the many pressures for translators nowadays in terms of turn-around time, rates, etc., the advantages of tools such as machine translation and translation memory should be obvious.

The workshop then delved deeper into the subject of the steps required to implement the use of MT for freelance translators, discussing the pros and cons of using MT and the advantages of MT over TM.

The internet has become a wonderful tool to survey the available programs and to choose the ones that most likely would be suitable for the professional translator. For hands-on experience, the participants had to evaluate the output of three providers of free MT.

The next step in the selection process provided a brief introduction into pragmatic evaluation and also involved some hands-on practice. Again, it was a comparison of source text processed by three professional-grade MT programs, focusing this time on the relative ease or difficulty of converting the raw MT into publication-ready text.

Finally, the workshop focused on a step-by-step guide through the development of an MT project. All the necessary steps for successful MT use, from previewing the source text to pre-editing, lexical work, post-editing and archival work for TM were reviewed. To get a feel for the potential for time and effort savings, this process was compared to the general manual translation process, and it could be shown that there is a definite advantage in the use of MT once one begins follow-up projects. This, of course, was cause for lively discussion that went on well beyond the official end of the workshop. Unfortunately, this left no time for the last part of the agenda, the quick course in post-editing. Overall, however, it seemed that the workshop was well received.

Walter Hartmann, MTConsulting Co. wh@mtconsult.com ❏

SDL Lands Clients

service will be expanded to include more languages, including Dutch, during the next 12 months.

The other client is Trias Politica BV, a newly created subsidiary of Compendium Group BV in The Netherlands, which recently launched a European information service for corporate and government bodies. The site provides these organizations with vital information, articles and documents from the public sector and offers specific coverage of competition and tendering issues.

ETS will optimize the Compendium online subscription service for its multilingual audience, allowing for instant translations of document searches. SDL's ETS was chosen for its translation speed, its ease of integration with the portal search technology and its dictionary management functionality - allowing the import of Compendium's rich thesauri and glossaries. The online translation service will support French and German to English translations in a first stage, then expand to cover all European languages as the service is rolled out to a larger multi-language audience. SDL will create Trias Politica specific dictionaries from the text material and thesauri available. Compendium realise that some users will expect a 'perfect' translation of a legal document for further publishing and so SDL will also provide an online link to their human translation service, enabling a quick and professional localisation service for selected users. ❏

Sakhr

required, and the text sentences or object images are transparently localized. Texts are automatically translated, while human translators guarantee 100% accuracy. And the real advantage is that the customer doesn't have to build an Arabic site because the corresponding Arabic site is generated on the fly".

Sakhr is a developer of software for professional users of Arabic, and operates one of the region's largest translation services. Digitek International represents Sakhr in the United States and Canada. *Contact Mark Meinke at Digitek International, +1-703-883-0134.* ❏

Oki

Management of community dictionaries is flexible. The creator of a community can serve as the leader and manage the dictionary individually, or all registered members can participate in the management process of adding new words, updating translations, etc. Other convenient features include: glossaries that users can post for translation by other members of the community, bilingual chatting supported by machine translation, and related links selected by each community. Translations can be produced from webpages, formatted files and simple text. Translations performed on the website's top page (i.e. not within a community) use a basic dictionary, while those created within a community domain use both the basic and community dictionaries. The translation engine, which is written 100% in Java, can be used as a module in other application systems, such as, information-retrieval systems.

Contact: Ms. Naomi Takeuchi, Oki Electric Industry Co., Ltd. Tel: +81 3 3580 8950 Fax: +81 3 3581 5522 or press@oki.co.jp ❏

Multilingual Support

multilingual CRM tool and its associated chat feature, the company now plans to release the multilingual chat system independently as stand alone product. This service would include the AI wrapper that improves translation quality, and could effectively be used in the context of any client interaction.

Paramon is a 3-year-old company with about 25 employees. Paramon management plans to keep the company small to maintain its "think tank" atmosphere. The company's core technology has so far been released in two products, the ParaCRM™ system, and Aperture™, a business decision support system, though Aityan claims others are in the works.

To see a demo of the dynamic multilingual capabilities of ParaCRM™, contact marketing@paramon.com. Website: www.paramon.com. ❏

Resources

Laurie's Links
ASP MT Providers

Somebody needs to do a study of the newly emerging MT service providers. They are not developers, but all seem to be drawing on the same intuition—that people would use MT more if it were a little more convenient. For example if the MT component were housed on a server so users didn't have to mess with installations and upgrades; and then if it were packaged with other handy tools like email, instant-messaging, and collaboration; and finally, if users could try the MT output and then instantly submit the original text for a human translation all in one seamless transaction. This is not the beginning and end of the services below, but all have implemented at least part of the model. Alis (www.alis.com) pioneered this market niche, and they have now been joined by WorldLingo (www.worldlingo.com), TransClick (www.transclick.com), TranslationWave (www.translationwave.com), Amikai (www.amikai.com) and probably others. Recently WorldLingo announced that their tools and services would also be offered by an Italian business portal called Giallo (www.giallo.it). The question in my mind has been, "Are they making any money?" It is not clear how the business community has responded to these offerings.

Search tools

I tend to settle into using one Internet search tool until a colleague tells me about something cooler or faster. First it was AltaVista. Then Google stole my heart a year and a half ago. But there is a lot more to searching! For example there is a burgeoning site-search market, in addition to the well-known public and free Internet search engines. You can stay on top of the latest developments by periodically checking out www.searchtools.com. The tools list includes links and information for dozens of tools for every purpose, with handy annotations (Java, Perl, Unix, open source, etc.) There are also reports on conferences devoted to searching so you can find out what's still in the labs. ❏

So many events were announced, we couldn't fit them all in! See www.eamt.org/ mtni.html for additional events of interest to the MT community.

2002

January 8-9: CLUK—The Annual Computational Linguistics in the UK Colloquium. Guest Speakers: Mark Davies, Illinois State University and Jenny Thomas, University of Bangor. *See: www.comp.leeds.ac.uk/cluk5*

January 21-25: 1st International Wordnet Conference. Central Institute of Indian Languages, Mysore (Karnataka State), India. Organized by the Global Wordnet Association together with CIIL Mysore, India. *See: www.hum.uva.nl/~ewn/gwa. htm*

January 28-February 1: 20th International Unicode Conference. Washington, DC, USA Organized by the Unicode Consortium. See: *www.unicode. org/iuc/iuc20*

January 31-February 2: PACLIC 16. The 16th Pacific Asia Conference on Language, Information and Computation. Jeju Island, Korea. Hosted by the Korean Society of Language and Information. See: *www.paclic.org*

February 17- 23: CICLing-2002. Mexico City, Mexico. Second International Conference on Intelligent Text Processing and Computational Linguistics. *See: www.cicling.org/2002.*

February 28 - March 2: II International Conference on Specialized Translation. Barcelona, Spain. *See: www.upf.es/dtf/ activitats/2icst/2ctradsp.htm.*

March 1-3: ICON 2002 (Indian Conference on Natural Language Processing). Channai, India. *See www. iiit.net/icon_conf.htm*

March 13-17: TMI 2002. The 9th Conference on Theoretical and Methodological Issues in Machine Translation. Keihanna, Japan. See: *www.kecl.ntt.co. jp/events/tmi/*

March 29-30: HKTerm. Hong Kong. workshop on terminology in education, research, and applications. See: *cpct92. cityu.edu.hk/TAHK/index.htm*

March 24-27: HLT 2002. San Diego, California. Sponsored by DARPA, NSF, ARDA et. al. Organizers: Mitch

Marcus, University of Pennsylvania (general chair), David Yarowsky, Johns Hopkins University (co-chair). **Extended abstract submission deadline: January 7, 2002.** *See: hlt2002.org.*

April 24-25: Language Technology for Business Information Systems. Posnan, Poland. Organizers: Jakub Piskorski, DFKI GmbH Germany, *piskorsk@dfki. de,* Feiyu Xu, DFKI GmbH Germany, *feiyu@dfki.de.* **Submission deadline: December 15, 2001.** *See: bis.kie.ae. poznan.pl*

May 9-11: The Fifth Symposium on Natural Language Processing 2002 + Oriental COCOSDA Workshop 2002. Hua Hin, Prachuapkirikhan, Thailand. **Submission Deadline: February 1, 2002.** *See: kind.siit.tu.ac.th/snlp-o-cocosda2002*

June 24-27: TALN 2002. Organized by l'ATALA (Association pour le Traitement Automatique des Langues). Nancy, France. **Deadline for tutorial proposals: January 31, 2002. Deadline for paper submission: February 15, 2002.** See: *www.loria.fr/projets/ TALN*

June 10-12: ISWC—The International Semantic Web Conference. Sardinia, Italy. **Tutorial proposal deadline: January 21, 2002; Paper submission deadline: February 15, 2002.** *See: iswc.semanticweb.org.*

June 24-26: Recital 2002—Student Sessoin at TALN. Nancy, France. **Deadline for paper submission: February 15, 2002.** See: *www.loria.fr/projets/ TALN*

July 1-3: INLG 2002 (Second International Natural Language Generation Conference). **Deadline for paper submission: February 14, 2002.** See: *www.research.att.com/~rambow/ inlg/inlg.html*

July 7-12: ACL-2002. University of Pennsylvania, Philadelphia, PA, USA. Organizers: Pierre Isabelle (general chair), Eugene Chainiak and Dekang Lin (program co-chairs) Martha Palmer

(Local Arrangements). **Deadline to submit tutorial proposals: January 11, 2002. Deadline for paper registration: January 25, 2002.** *See: www. acl02.org*

July 15-26: ELSNET Summer School. Odense, Denmark. **Pre-registration deadline: April 15, 2002.** *See www. elsnet.org, or contact Holmer Hemsen hemsen@nis.sdu.dk for more information.*

August 7-10: 16th World Congress of the International Federation of Translators, "New Ideas for a New Century" Submission deadline, September 1, 2001. *See:www.stibc.org/fit2ki.htm.*

August 24-September 1: COLING 2002. Academia Sinica, Taipei, Taiwan. Organizers: Chu-Ren Huang, Academia Sinica (Conference Chair); Winfried Lenders (Program Chair), Univ. of Bonn, Germany *Lenders@uni-bonn.de;* Antonio Zampolli (Workshop Chair) *pisa@ilc.pi.cnr.it.* **Deadline for Workshop Proposals: January 15, 2002 Deadline for paper submission: February 15, 2002.** *See: www.coling2002. sinica.edu.tw*

Sepbember 18-20: DAARC2002. Discourse and Anaphora Resolution Conference. Lisbon, Portugal. Organisers: Antonio Branco, Tony McEnery and Ruslan Mitkov **Submission Deadline: January 4, 2002.** *Contact: daarc2002@di.fc.ul.pt*

October 8-12: AMTA 2002. Tiburon, California, "From Research to Real Users". Organizers: Elliott Macklovich, University of Montreal (General Chair) *macklovi@iro.umontreal.ca;* Stephen Richardson, Microsoft, (Program Chair) *steveri@microsoft.com;* Violetta Cavalli-Sforza, University of California at San Francisco (Local Arrangements) *vcs@sfsu.edu;* Robert Frederking, Carnegie Mellon University (Workshops, Tutorials, and Website) *ref@cs.cmu. edu;* Laurie Gerber (Exhibits) *lgerber@gerbersite.com.* **Submission deadline: April 15, 2002.** *See www. amtaweb.org for latest developments.*