



MT News International

Newsletter of the International Association for Machine Translation

In this issue . . .

Spotlight on the News	1
Products/Announcements. . .	3
Conference Reports	6
Features	8
Conference Announcements	14

Of special note ...

- **MT Summit IX
Announcement** (page 14)
- **Extensive 2002 Conference
Reports** (starts page 6)
- **Jaap van der Meer on
Translation Automation**
(starts page 8)
- **Feature Article: "From the
Garage to the Attic" by
David Clements** (pages 10-11)
- **Upstart Data-Driven MT
Companies** (pages 18-19)

Spotlight on the News

EAMT-7/CLAW-4 Joint Conference

*Dublin City University, Ireland
May 15-17, 2003*

The theme of the the 2003 joint conference of the European Association for Machine Translation (EAMT) and the Controlled Language Applications Workshop (CLAW) is "controlled translation". Papers addressing this theme will be featured on the second day of the conference, with the first day devoted to general papers on machine translation (MT), and the final day dedicated to other papers focusing more on controlled language issues.

Background

Over the years, there have been many conferences on MT, involving rule-based approaches, statistical and example-based approaches, hybrid and multi-engine approaches as well as those limited to particular sublanguage domains. In addition, there has been an increased level of interest in controlled languages,

7th International EAMT Workshop at EACL'03

*Budapest, Hungary
April 13, 2003*

There is an ever-growing need for tools for translation, and Europe is envisaging a particular challenge with the enlargement of the European Union which will add 10-12 new languages.

One of the important trends emerging from the 2001 MT Summit in Santiago de Compostela was that MT is going more and more towards an integration or combination with other tools.

At this workshop we want to follow up on these two trends by focussing on how MT and other language technology tools can be combined in order to produce translation faster and better, and on how language technology tools can support faster production of MT systems.

Theme 1: Various ways of combining 'proper' MT with other types of language technology tools in order to improve performance and efficiency of translation have been discussed. This may in-

Continued on page 2 ►

Continued on page 2 ►



MT News International

Issue No. 32 (vol. 11, no. 2)
Spring 2003

EDITOR-IN-CHIEF:

Laurie Gerber

Tel: +1 (619) 200-8344
Fax: +1 (619) 226-6472
E-mail: mtni@eamt.org

CONSULTING EDITOR:

John Hutchins

E-mail: info@eamt.org

CONTRIBUTING EDITOR:

Colin Brace

Fax: +31 (20) 685-4300
E-mail: webmaster@eamt.org

REGIONAL EDITOR, AAMT

Hitoshi Isahara

Fax: +81-774-95-2429
E-mail: isahara@crl.go.jp

REGIONAL EDITOR, AMTA

David Clements

E-mail: dclemen1@san.rr.com

REGIONAL EDITOR, EAMT

Jörg Schütz

Fax: +49 (681) 389-5140
E-mail: joerg@iai.uni-sb.de

Copyright 2003 by IAMT. Permission is hereby granted to reproduce articles herein, provided that they appear in full, and are accompanied by the following notice:

"Copyright 2003 International Association for Machine Translation (IAMT). Reprinted, with permission, from the IAMT newsletter, Machine Translation News International, issue #32, March, 2003."

Electronic copies available upon request: mtni@eamt.org

EAMT/CLAW

...continued from page 1

culminating in the series of workshops on controlled language applications (CLAW). These have given impetus to both monolingual and multilingual guidelines and applications for using controlled language for many different languages.

Controlled languages are subsets of natural languages whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity. Traditionally, controlled languages fall into two major categories: those that improve readability for human readers, particularly non-native speakers, and those that improve computational processing of the text. It is often claimed that machine-oriented controlled language should be of particular benefit when it comes to the use of translation tools (including machine translation, translation memory, multilingual terminology tools etc.).

Experience has shown that high quality MT systems can be designed for specialized domains (e.g. METEO). However, the area of controlled translation has remained relatively unaddressed. This is rather strange given its undoubted importance. Such examples that exist use rule-based MT (RBMT) systems to translate controlled language documentation, e.g. Caterpillar's CTE and CMU's KANT system, and General Motors CASL and Lant-Mark, etc. However, fine-tuning general systems designed for use with unrestricted texts to derive specific, restricted applications is complex and expensive.

There are several examples of using Translation Memory (TM) tools in a controlled language workflow, yet these have been primarily for combining TM and MT tools. Very few attempts have been made where Example-based MT (EBMT) systems have been designed specifically for controlled language applications and use. This is even harder to fathom: using traditional RBMT systems leads to the well-known 'knowledge acquisition bottleneck', which can be overcome by using corpus-based MT technology. Furthermore, the quality of EBMT (and Translation Memory) systems depends on the

Continued on page 13 ►

EAMT/EACL

...continued from page 1

clude pre-editing tools, taggers, post-editing tools, access to bilingual concordances, term extraction tools, categorisation tools, semantic clustering etc. The workshop will briefly cover state-of the art and attempt to discover new directions for improving MT by embedding it in an environment of other tools.

Theme 2: The growing demand for MT for new language pairs makes it necessary to find ways of supporting the production of new language pairs. Both providers and researchers are using resources, statistics, and language technology to make progress in this field. The second half of the workshop will concentrate on these efforts and their success or failure.

The workshop will end with a panel discussion on trends and future perspectives.

This is the first time an EAMT workshop is organised at EACL and we hope to have combined, in a very integrated way, the interests of the two communities, the EAMT community and the EACL community, so we are looking forward to a very successful workshop!

Programme Committee:

Bente Maegaard, CST, Copenhagen, DK (chair)

Harold Somers, UMIST, Manchester, UK

Jörg Schütz, IAI, Saarbrücken, DE

Gábor Prószéky, Morphologic, Budapest, HU

Contact for Submissions and Submission Inquiries

Bente Maegaard, Center for Sprogteknologi, Njalsgade 80, DK-2300 Copenhagen S, Denmark

Phone +45 35 32 90 90, Fax +45 35 32 90 89

E-mail: bente@cst.dk



Word Magic Software Launches in the U.S.

[adapted from press release]

Word Magic Software, a developer of English-Spanish translation and communications software, launched products in the U.S.A. in 2002. The Word Magic's current translation products include: English-Spanish Interpreter 3.14 (ESI), Translation Dictionary & Tools, Field-Specific Dictionaries (Legal, Business and Finance, Information Technology) and the Point and Click Translator. All products are available for both Spanish->English and English->Spanish.

According to the U.S. Hispanic Chamber of Commerce, 45% of U.S. trade is taking place in Latin America and the Caribbean, with most Fortune 500 companies having operations in this region as well. Given the increased demand for bilingual communications, Word Magic Software, is focused on penetrating the U.S. marketplace and securing its position as the specialist in the Global Translation and Communications Industry for English-Spanish communications. Word Magic Software is focused on providing companies and consumers with the tools necessary to communicate with increased precision and efficiency across Spanish and English language frontiers, while at the same time aggressively continuing research and development initiatives in Machine Translation and Artificial Intelligence.

Special features of Word Magic translation products are:

ESI: Text to Speech output, grammar display, spell checker, Spanish ambiguity checker, and an interactive translation mode that allows users to select the correct meaning in context.

Point and Click Translator: Allows users to right-click on any word in any active window, and get access to the word's translation, synonyms, and conjugations. A double-click replaces the word with its translation.

Dictionary resources: Over 800,000 translation references, 225,000 idiomatic references and 4 million synonyms are available.

Word Magic Software was incorpo-

rated in 1989 and is headquartered in San José, Costa Rica. For more information, see: www.wordmagicsoft.com. Contact: Maria Bacallao/Denise Galvez, Anthony Baradat Iglesias Advertising & Public Relations 305-859-898; Mbacallao@abiadvertising.com; Dgalvez@abiadvertising.com □

HLTheses Web Site Announced

[adapted from press release]

HLTheses is a web-site devoted to current research in Human Language Technology (HLT). It is an ELSNET initiative, co-sponsored by ISCA and EACL, aiming to aid HLT researchers and promote their research by providing information on relevant PhDs worldwide. HLTheses contains extended PhD abstracts (completed and on-going), CVs and contact information of PhD authors and researchers. The site intends to reference all PhD theses related to HLT worldwide, thus serving as a state-of-the-art guide and a place where researchers can find others who work in similar areas.

Call for PhD Abstracts

HLTheses invites all researchers holding or pursuing a PhD in any area of Human Language Technology (HLT) to submit extended abstracts of their PhD theses (and brief CVs) in order to be listed in the HLTheses repository. The PhDs need not be completed. On-going PhD abstracts are equally eligible. PhDs presented in HLTheses are expected to create a great impact, since HLTheses will initiate a large promotional campaign. HLTheses intends to provide additional annual printed volumes containing the theses completed every year, sorted by research area.

See: hltheses.elsnet.org.

Contact: Nikos Fakotakis, HLTheses Editor-in-Chief, fakotaki@wcl.ee.upatras.gr □

Cyasoftware Offers Middle Eastern Languages

[adapted from press release]

Cyasoftware, headquartered in Virginia in the U.S.A. offers transfer-based machine translation systems for some of the languages of greatest interest in the current middle-eastern conflict: translation systems for English to and from Farsi, Dari, and Pashto. In addition, the company offers a number of other products and services, including: an OCR for Farsi, Dari, Arabic and Urdu; text to speech (TTS) software for Dari, Farsi, Pashto and Arabic; automatic speech recognition (ASR) software for Dari, Farsi, Pashto, Urdu and Arabic; as well as custom software development and human translation services for middle-eastern languages.

See www.ciyasoftware.com or contact Naquib Hatami, info@ciyasoftware.com. □

InstantService Adds MT to Call Centers

In MTNI No. 29 (December 2001) we reported on a California Company, Paramon, that had recently begun offering customer support tools incorporating machine translation. That offering is still available, together with much more information and a growing suite of products at the company's increasingly sophisticated website: www.paramon.com

In 2002, another company, Instant Service, headquartered in Seattle, Washington, added machine translation to its contact center offering. Instant Service offers hosted solutions for contact centers and customer support for small to medium businesses. InstantService solutions allow companies to quickly get up and running with Web-based chat, email, and FAQ lists. The fact that the service is hosted by InstantService means that no additional infrastructure is required to get started. Having such tech-

Continued on page 13 ►

Lexicography Courses Offered at University of Brighton

As our understanding of effective machine translation deployment grows over time, it becomes clear that good lexical development and customization is key to improving translation quality. However, so far, there have been few academic courses focused on lexicography, particularly for computational linguists, and thus it has been left to the determined translator, linguist, programmer etc. to learn the lessons of lexicography for themselves. Beginning in the current academic year, the University of Brighton's Information Technology Research Institute (ITRI) is offering a Master of Science in Lexical Computing and Lexicography, as well as shorter focused courses of study that can fill this gap.

The MSc coursework consists of ten modules, seven compulsory, and three elective. Eight modules are required to complete the program. The individual modules of instruction are also open to qualified students from outside the university on an a la carte basis. The individual modules provide an opportunity for working lexicographers, computational linguists and language teachers to get useful training to advance their careers. Likewise, employers may benefit from the opportunity to bring in more current skills and methods for lexicon and terminology development in translation work and NLP system development.

The courses at Brighton are taught by an experienced team of lexicographers and computational linguists who form the core of the regular research staff at ITRI: Sue Atkins, who has edited and guided the development of a number of significant published dictionaries, including Collins-Robert and Oxford-Hachette English-French dictionaries. Ms. Atkins was a founder member and is a Past President of the European Association for Lexicography and was the driving force behind the British National Corpus. She is currently developing, alongside Charles Fillmore, the Frame

Semantics paradigm. Adam Kilgarriff worked as Longman Dictionaries Computational Linguist prior to joining ITRI in 1995. He is a leading authority on the interface between lexicography and language technology and has published extensively in both areas. A paper on the lexicographers' workstation, developed by the lexicography research group at ITRI, appears in the proceedings for MT Summit VIII, held in Santiago de Compostela, Spain in September 2001. Further research will be reported at the upcoming European Association of Computational Linguistics Meeting to be held in Budapest, Hungary, April, 2003.

The MSc program, currently in its first year, has eight students, primarily lexicographers and linguists looking to develop current computational skills. Additionally some students are computer scientists looking to develop skills for lexicography and NLP work. Students enrolling a la carte in the course modules tend to come from industry: lexicographers from publishing companies, and technical people from computing companies.

The vision for the program is based in the idea that lexicography and NLP need each other: NLP needs good lexicons as a resource. Lexicography, even for relatively conventional published dictionaries, can benefit tremendously from contemporary methods which can increasingly do much of the work of collecting examples automatically, as well as extracting useful grammatical information from texts to establish patterns of usage. One of the interesting developments from the research group at ITRI has been "word sketches," which are lexical entries automatically extracted from text using *mutual information* statistics, and which capture very rich information about word behavior, including generalizations about the common subjects and objects of verbs. This kind of information seems like it could greatly assist developers of MT systems, particularly of transfer systems, which generally rely on linguists' observations of contextual examples to encode this information in dictionaries. The problem in such endeavors is always the limitations on the amount of text a lexicographer can effectively review and digest. This limita-

tion is removed when the data can be summarized automatically.

While the focus of this article has been on the usefulness of such training for development of dictionaries for MT, in fact, the courses and experience of the instructors provide tremendous insight into the business of publishing dictionaries as well, including issues such as the software used for development and typesetting.

The course modules currently taught are:

Core curriculum: Lexicography 1; Lexicography 2 (double module); Bilingual Lexicography and the use of Parallel Corpora for Translation; Linguistics and the Lexicon; Lexical Computing 1; Lexical Computing 2: enriching the corpus (double module); Corpus Design and Use

Electives: Lexical Computing 3: automating lexicographic tasks; Dictionary Project Management; Corpora and Language Teaching

For more information, see the ITRI website: www.itri.bton.ac.uk/courses. □

Lingvistica Unveils Resources for Developers

[adapted from press release]

In addition to full-fledged machine translation systems, Lingvistica '98 Inc. and Lingvistica b.v. have begun offering Grammatical Dictionaries for European and Asian Languages. The goal is to develop and supply to the international market with grammatical dictionaries as text files to be used in various language engineering projects such as machine translation, automatic abstracting and indexing, web mining, search engines, etc.

Each grammatical dictionary includes words (not less than 100,000 for each language) and morphological features, such as POS, declension/conjugation models, etc.

The first grammatical dictionary in this project developed by Lingvistica '98 is that for the Polish language. It in-

cludes almost 120,000 words with grammatical features (as well as pronunciations) attached to each word.

Monolingual dictionaries have been developed and are being tested and proof-edited for German, Russian and Ukrainian. Additional dictionaries are under development for Dutch, English, French, and Turkish. In addition, bi-directional dictionaries are being developed as linguistic resources.

In this project, Lingvistica is working in tight collaboration with LogoMedia, the world-known Boston-based MT developer, SCIPER, a French company specializing in developing linguistic resources, The CJK Dictionary Institute, Inc., based in Japan, and Onyx Consulting and Computer Research Laboratory, New Mexico, USA.

See: www.ling98.com and www.lingvistica.com.

Contact: Dr. Michael S. Blekhan, President, Lingvistica '98 Inc. And Lingvistica b.v. tel: (514) 331-0172; ling98@canada.com.



New Masters Course Available at U. of Cambridge

[adapted from press release]

A new masters course has replaced the highly successful M.Phil in Computer Speech and Language Processing at the University of Cambridge.

Like its predecessor, a key aim of the masters course in Computer Speech, Text and Internet Technology is to teach the fundamental theory of speech and natural language processing. However, the new course also focuses on its application to information management and access within the framework of emerging Internet and W3C standards, such as XML text and speech annotation.

The one-year graduate course runs from early October to end of July and consists of two terms of lectures and practicals followed by a three-month project. The final degree is awarded on

the basis of coursework, examination and project. Studentships are available.

The course differs from some other programs by providing an in-depth practical and theoretical grounding in the techniques for speech and language processing, which form the basis for today's commercial and research prototype systems. There are strong links with industry and many of our past students have gone on to work for high-tech start-ups and industrial research laboratories, either immediately or after completing a PhD.

To further strengthen our links with industry, the course is also available to students wishing to pursue it on a part-time basis. Cambridge is a major international centre for research in both speech and language processing. The course is taught by leading researchers in these areas who have active collaborations with industrial and academic laboratories in Europe, the US and Japan.

See: www.cl.cam.ac.uk/Teaching/CSTIT/ or contact: Lise Gough, University of Cambridge Computer Laboratory, Cambridge, UK. Tel: +44 (0) 1223 334656; cstit-enquiries@cl.cam.ac.uk.



LTC and CAS Integrate MT and TM with Automated Management Facilities

[adapted from press release]

The Language Technology Centre, in conjunction with German software developer CAS Software as technology partner, and two user organizations in Italy and Greece obtained funding from the eContent program of the European Commission for a trial. The consortium is in the process of developing a web standard compliant interface and testing the integration of translation technology and customized machine translation within a well defined domain and user environment.

A first pilot was ready at the end of June 2002 and is currently being tested and linguistically fine tuned. CAS Software and LTC as technology partners intend to release the outcome of the project as a product

suitable in a variety of multilingual web communication environments.

About LTC

Language Technology Centre (LTC) was established in 1992 by Dr Adriane Rinsche with the objective of providing language technology solutions to a wide variety of potential application areas. LTC specializes in building multilingual websites, software localization, consultancy in language technology, technical translation, and software development.

Contact: Dr Adriane Rinsche, The Language Technology Centre Ltd, Surrey, UK: +44-20-8549-2359; info@langtech.co.uk. See : www.langtech.co.uk



Latest Online MT Compendium Now Available

The latest edition of the "Compendium of translation software" (March 2003) is now available on the EAMT website: <http://www.eamt.org/compendium.html>.

If you see any errors or omissions please report them to John Hutchins. He promises to include changes in the next update due in August.

For those interested in seeing what was available in the past, there are back editions in PDF format available on Hutchins's website:

<http://ourworld.compuserve.com/homepages/WJHutchins>.

Contact: John Hutchins wjhutchins@compuserve.com.



Conference Reports

JEITA Symposium, October 2002

By Jonathan Lewis

At a JEITA symposium held in Tokyo on October 16, 2002, Dr. Edward Hovy of the Information Sciences Institute, University of South California, gave a lecture entitled "Positioning Natural Language and Speech Processing Research to Exploit New Opportunities: Trends in North America and Europe." Professor Hovy addressed three questions in his lecture: Where are we in language processing now? Where will we go soon? And where is the money?

He started with a modest success story. In 1999 the US government was very worried about the situation in Indonesia, but it lacked Indonesian speakers able to gather information. So DARPA asked for help from NLP researchers, who put together a system in only three months. They built a web spider for Indonesian, which returned results in similar fashion to Google. A summary was produced in Indonesian, then an English translation of the summary. The system only worked for Indonesian, could not be operated by voice, and the system didn't learn. Nevertheless, the fact that it was used for a year illustrated the point that "even if the components are not very good, you can put them together to get 'good enough' applications."

Dr. Hovy then took the audience at breakneck speed through the fifty-year history of NLP, information retrieval (IR) and speech processing. In the first decades (roughly 1950-1965) researchers in all fields tried out a whole range of techniques, with only limited success. In the second decade (1965-1975), while NLP researchers struggled to apply Noam Chomsky's theoretical breakthroughs the significant progress was made in information retrieval, where researchers started to work on the vector space rather than the word level.

These technologies remain useful today.

In the third decade (1975-1985) the big change was in speech processing, which saw the triumphant introduction of learning methods. The fourth decade (1985-1995) finally saw the long-awaited leap forward in NLP, which came through work on statistical MT at IBM. The statistical breakthrough caused great progress to be made on the practical side, such as tools, and evaluation competitions. In the fifth decade (1995-2005) we are witnessing the triumph of statistical methods, with every paper making some use of evaluation, corpus, or large numbers.

The first lesson we learn from this history is that language processing is ENGINEERING. We don't have a deep theory of mind and syntax. Instead of trying to define the rules of syntax and semantics, we should get systems to learn the rules themselves. The second lesson is that language processing is NOTATION TRANSFORMATION. Each small transformational step can be a module, and these modules can be combined in new ways that together will be just good enough.

Today, NLP still cannot do general-purpose text generation. It can't deliver semantics. It can't deliver long/complex answers by extracting, merging and summarizing information on the web. It can't handle extended dialogues, or use pragmatics (style, emotion, user profile). And it can't contribute significantly to the theory of language.

On the other hand, we can achieve accuracy levels of between 40% and more than 90% in a whole range of tasks ranging from information retrieval to shallow parsing. The significant point is that all these accuracy rates jumped in the years after 1985 when we finally learned the lessons of modularization. We are now seeing a move towards automated knowledge acquisition, evaluation, rapid build-evaluate-test cycles, and toolkits. And we are moving away from deep processing, semantics, inference, parsing, and generation -- we are still struggling to get information into the system, rather than get information out.

Dr. Hovy then outlined short- and long-term research projects in the United States and Europe. He noted the great opportuni-

ties that the Web offers to developers of NLP applications. As far as the semantic web is concerned, the dream is of a Web in which each page is supported by a semantic (interlingual) description that is used by search engines and then translated by presentation engines into a useable format. But this vision will remain just a dream, due to the limitations of NLP and other technologies. However, a weaker version in which each text/speech Web page contains a limited set of annotations to be used by search and display engines is quite achievable.

Turning to the long-term payoffs of NLP, IR and speech technologies, Dr. Hovy stressed that we should break transformations down into small steps and create a toolkit of small transformers. This allows us to adapt quickly to new challenges, re-assembling small transformers in new ways. We can also try to hide the quality weaknesses of one module by the power of other transformer modules. To give one example, Hovy's colleague Kevin Knight developed a tool for name transliteration from Japanese katakana to English. The problem is complex, but solvable when one breaks it down into steps. Knight's system took 8 or 9 months to build, but it works better than people now.

Dr. Hovy saw obvious opportunities for application developers that would be perfectly usable with current technologies: For example, a handheld tourist assistant that could answer simple questions; a business watch service to inform the user of developments in the industry or market; an information gathering service for journalists or students. The question in each case is why aren't we already building these things?

Finally, Dr. Hovy asked how we are going to achieve better quality translation and more correct results. He saw a need to work in two areas. First, meaning. We have to get back to semantics, to develop ontologies, semantic interpretation and generation. This will be our big problem ten years from today. The way forward is to learn about these from the Web, through text-mining. Second, better interfaces, especially better dialogue systems and adaptive user models. In other words, current systems such as Web search engines don't know what the user already knows or has already been shown.

Jonathan Lewis is an Associate Professor at the Institute for the Study of Global Issues at Hitotsubashi University, Tokyo, Japan:
jonathan_lewis@mac.com



ASLIB Translating and the Computer 24

By Daniel Grasmick and Christopher Pyne, SAP AG

The 24th edition of the "Translating and the Computer" conference organized by ASLIB was held in London on November 21 and 22, 2002.

Over 150 delegates attended this conference and the organizers were delighted to see an increase in participants compared to 2001.

The keynote presentation on the first day was given by Hans Uzkoireit, Director of the DFKI in Saarbrücken, Germany. He reviewed the evolution of the various linguistic approaches applied especially in machine translation during the past years and discussed evolution as well as the current trends.

Daniel Gervais from MultiCorpora Inc., Canada, then presented their MultiTrans tool working on a full text multilingual corpus and claimed their superiority compared to traditional Translation Memory approaches.

Then Doug Lawrence from Trados jumped in to present Matthias Heyn's paper which focused on the value of linguistic assets and investigated the potential of upcoming technological innovations. He argued that language technology has to cater to the peculiarities of an inherently fragmented supply chain in order to meet the user needs.

To complete the morning sessions, Gábor Prószéky from MorphoLogic, Hungary, presented their pattern-based Machine Translation system from English to Hungarian.

After lunch D. Verne Morland from NCR Corporation, USA, described NCR's experience with the introduction of MT applied to its Personal Learning News, a monthly email newsletter.

Reinhard Schäler from the University of Limerick, Ireland, gave a stimulating overview of the various strategies developed by translation companies in order to survive the current economic climate.

Carl Kay, an independent experienced consultant from Japan, gave an introduction to the nature of the Japanese lan-

guage, and how that language affects the current state of the translation industry and language technology tools in that country.

A reception sponsored by CLS Corporate Language Services AG, Switzerland, concluded the very successful first day of the conference. Due to the sometimes controversial presentations and the high quality of the delegates, dozens of questions were asked and a lively exchange took place.

Day two covered many practical and some conceptual themes ranging from controlled authoring, handling tagged text formats in translation tools to evaluating term extraction software.

SAP's Christine Thielen presented the SKATE controlled authoring project, along with Dr. Andrew Bredenkamp, managing director of acrolinks, SAP's development partner. Both presenters stressed the very practical approach adopted by the project teams and how critical usability and flexibility issues were for end-users. The system, which is in the process of being rolled out, currently covers SAP's two authoring languages, German and English.

The morning continued on a practical note with the presentation by Lorna Joy on the issues posed on translators having to handle tagged text formats such as HTML and XML. Lorna made the important point that tagged formats were depriving translators of valuable context information and that tag manipulation was sadly becoming an important, if not particularly rewarding activity for translators. She went on to demonstrate how she avoided "Tag Tyranny" at her company, SCHÜCO.

The fairly recent commercial availability of term extraction tools and their increasing importance within the translation environment was a driving factor for Véronique Sauron from the University of Geneva in her presentation on how three such tools could be evaluated. In addition to some very interesting results from the first stage of the project, Véronique stressed that an important element was the establishment of a standardized methodology for evaluating such tools.

In a very entertaining presentation before lunch, Yves Champollion, the developer of the Wordfast translation memory system, put forward the idea that the future lay in combining of translation memory and machine translation approaches. Using ideas borrowed from the mathematical

Chaos Theory, Yves suggested that massive volumes of translation data could be made less redundant or "shrunk" to form structural matching databases which themselves can point to the correct content in an "ocean of (translation) data".

Mike Roche of IBM, Dublin and Monika Röthlisberger-Kaiser from CLS Switzerland presented interesting and different papers relating to translation services. Mike looked at the increasing role of the web in translation transactions and especially how one company, Berlitz collaborated with IBM to produce a web service for the submission and retrieval of translation jobs. Mike also discussed the importance of the XLIFF-based web services standard for translations.

Monika talked about the CLS workflow application for translation services. This Lotus Notes based system allows CLS to meet tight deadlines by ensure that all actors in the system do not lose time through file handoffs or poor information flow. Monika reported that the system was working very effectively and that new modules for accounting and other services were being added.

Dan Dube from the Isogen Corporation looked at some of the issues involved in authoring and producing information in XML. Dan presented the notion that with the correct environment, use of business rules to manipulate and streamline XML structures could help to achieve significant savings in translation time and effort. Finally Dan turned to formatting and DTP issues and showed how some of these could be overcome using XLSFO (XSL Formatting Objects), even for complex Asian and bi-directional layouts.

Finally, Lee Guillam from the University of Surrey rounded off the proceedings with an entertaining and thought-provoking insight into the use and importance of standards in language and knowledge engineering systems.

The organizers and the program committee were confirmed in their perception that the mixture of technology, use cases and theoretical presentations was very well accepted.

We look forward to ASLIB T&C 25. For more information, see www.aslib.com/conferences.

□

Translation Automation Finally Achieved: An Overview of the Translation Market

By Jaap van der Meer

Big changes are about to happen in the localization industry that will overwhelm the capabilities and expectations of the current suppliers, large and small.

Fifty years after research on machine translation commenced, widespread translation automation finally seems about to become a reality. The deciding factor is perhaps less the improvements to the technology than the insurmountable need for translated content. The current shape of demand has been formed by the shifting requirements through the years, and is easiest to understand when viewed as a composite of the three phases in the history of the translation industry: the translation phase, the localization phase and the globalization phase. First we look at the three phases and then explore what globalization means for the organizations undertaking it, the GIL (Globalization, Internationalization and Localization) industry that serves them, and the translation technologies that can make the challenges manageable.

Translation Phase

The translation phase started with the industrialization of translation services in **the nineteen fifties**. The first washing machines, radios and television sets required multilingual booklets. Agencies were set up by free lance translators, who often were political refugees after the turbulent times of the Second World War and its aftermath. Agencies would work on user instructions, "overseas" trading contracts and the translation of diplomas for the flows of emigrants to the New World. Translation service was not very time-critical. The means of delivery was regular post or telex. The primary concern for clients was cost. Judging from the multilingual booklets we still get with our home appliances, quality was and is of the least concern to the publishers. After all, who would really look into the user instructions for a refrigerator? And if they did, the only concern was really whether

the user would be able to find the switch for defrosting the freezer. The translation market blossomed and many new agencies were formed in **the sixties and seventies**. They all remained relatively small and serviced a close group of local customers. The main means of business acquisition was through yellow page advertising. Clients would prefer to work with an agency around the corner, even if they would hardly ever meet the owner of the agency face to face.

Localization Phase

The arrival of desktop computers and end-user software products marked the first big change in the translation market. In **the mid-eighties** new industrial clients emerged from the software industry. Their user instructions were not just small booklets, but big manuals of one thousand pages or more. The thing that complicated the matter was that there was not really a conventional product that went along with the manual. In fact the manual was the product. Consumers were not used to the new concept of "software", a product that you could not really see or touch. The manual therefore became the most important part of the software product. Yet another complication was that a good part of the information that required translation was embedded in the software product. User instructions and messages would appear on the screen of the computer. And obviously these texts should be fully consistent with the translated text in the manuals. It is no surprise that new specialized translation companies were established who focused in on the requirements of these new clients from the software industry. The new players took special care that their skills would not be confused with ordinary translation services: they called them-

Translation quality was hardly important for the manufacturer of television sets and lawn mowers, but now for software publishers quality was of the utmost importance.

selves *localization* companies. A few existing translation agencies successfully converted themselves into localization companies, but most of the successful localization companies were upstarts in **the late eighties and early nineties**. They grew much faster than the first generation translation agencies because the work volumes were so much bigger in the software sector. Their activities expanded outside the pure translation activity. They acquired skills to do complete page lay-out, graphics and also to perform testing and engineering work to ensure that the translated products would still function properly. As the volume of work grew bigger and the activities became more complex, they also needed project management skills to monitor project progress and timely delivery.

The big contrast from a customer perspective between the translation phase and the localization phase was the emphasis on quality. Translation quality was hardly important for the manufacturer of television sets and lawn mowers, but now for software publishers quality was of the

utmost importance. Consumers would not buy the software product if the manual was poorly written. The primary motivator for the software publishers to localize was to generate additional

revenues with the products that were already developed for their home markets. A badly localized product would get a bad review and would not sell very well. Software publishers would carefully select their target markets and calculate their return on investment based on the number of products they could sell in the German, French, Japanese or other foreign language markets. The dependence on quality caused a higher concentration of services with fewer providers. Customers would worry less about the vicinity of the agency to their own offices and more about the quality reputation of the company they would work with. Also, customers were concerned about the build-up of overhead in their own organizations to manage localization activities, which after all was still considered non-core business. The localization industry

went through a phase of mergers and acquisitions in the late nineties as a result of these market forces. However at the same time two other developments seemed to obstruct the road to prosperity for the few big and global localization companies. First of all localization skills were not so unique anymore: many small agencies had become very computer and localization literate, and they were competing with the large established localization companies. Second we saw a shift in the buying criteria from customers. Cost became more important than quality. Software publishers do not ship manuals that often anymore. Consumers are now used to the concept of software and they are happy to download the software. Instead of consulting manuals they prefer to call a call center or check a web site to find an answer to their query.

Globalization Phase

The beginning of the new millennium also marked the beginning of the globalization phase for the translation and localization market. The pervasive global presence of the Internet has caused a big change in the demand for localization work. Overall market share becomes more important now than individual product revenue. And market share is dictated not only by customer loyalty, but also by employee and supplier loyalty. Customers, employees and suppliers will feel more inclined to support and promote corporations with which they can identify themselves, companies that speak *their language*. A deep level of localization is required that can only be achieved if the localization process is really embedded in all enterprise processes. More and more companies are starting to realize this.

Historically, localization has been a standalone process, quite isolated from anything else that was happening in the corporation. For instance terminology that was translated by the localization department was not being shared by the marketing department or the training people. Integrating localization into the enterprise means that the IT managers (CTO-level) are (or should be) getting involved in localization of enterprise-wide systems. This expansion of the target for localization impacts the volume of text and timeframes. The volume of text to be translated multiplies as whole segments of the organization are localized, and time-

frames contract as companies struggle to get products to market before their international and in-country competitors.

These volume and time factors are driven by emerging trends which are evolving into standard practice: For these emerging global operations, it is no longer sufficient to translate product documentation and localize software, -- all enterprise information must be adapted to the global marketplaces in which the company wants to be seen as a market leader. Current and future clients learn everything about the company and its products through the company Extranet. They buy online, get support and get training online; they may even become members of company loyalty programs. At the same time, the company employs workers from many different countries who connect through an Intranet. They need to speak with "one voice" to customers. They need to be trained and supported in representing the company's products and services, and the company's distinctive philosophy. With market share as the single biggest goal, corporate branding becomes much more important. And finally, the company becomes more and more dependent on suppliers as outsourcing seems to be the way to go for most corporations. The suppliers are also spread out around the world. This dependence on suppliers creates a more intimate relationship, to the extent that it is quite likely the company maintains a secure web site for supplier management.

So far the translation and localization industry has been dealing with product documentation only. Product documentation has consisted of pre-sales materials

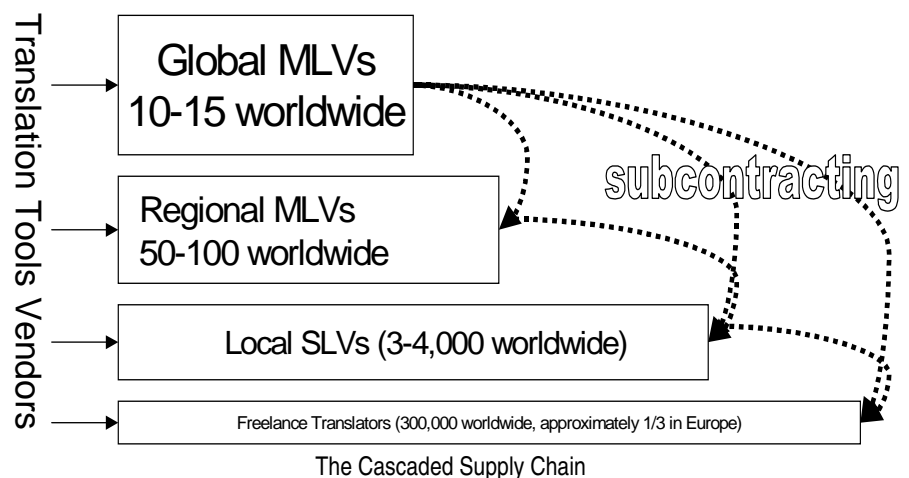
(brochures and advertising) and after sales materials (user documentation, software user interface, dealer information). In the globalization phase, the volume of material to be localized will grow to insurmountable levels. Companies will need to maintain multilingual versions of their knowledge bases, training materials, reports, proposals, human resource documents, etc. The product documentation probably represents less than 20% of all the text based information that will be subject to localization in the new globalization phase.

The other change is the time factor. Customers' needs are instantaneous. The quicker localized content can be made available, the more chance the company has to win new customers and retain existing customers. The translation and localization industry is facing an impossible task in this new globalization phase. Five times the volume must be translated in practically no time. In addition to the skills already built up during the localization phase, localization service providers are now required to be experts in XML, content management integration and globalization management software -- tools and standards that have emerged along with the globalization trend - to cope with the time pressure, and increased volume and complexity of translation requirements.

Cascaded Supply Chain

The evolution of the translation and localization industry through the three phases described above has lead to a

Continued on page 12 ►



Feature Article: From the Garage to the Attic

An Insider's View of Entrepreneurial MT

By David Clements

The Lernout & Hauspie saga first brought welcome attention, and then unwelcome attention, to the machine translation and language technology world. While most of us looked on from the sidelines, we have not heard much from those who were inside the rise and fall. David Clements' story kicks off a first person account that will be presented as a multi-part series. It takes us from the high spirited entrepreneurship that started MicroTac in San Diego, through the successive acquisitions and reorganizations involving Globalink, Lernout and Hauspie, Mendez Translations, and finally Bowne Global Solutions. David Clements, a veteran MT developer, is also the AMTA regional editor of MTNI. -Ed

Part 1: Reinventing the Wheel

I never intended to start a career in Natural Language Processing, nor in its peculiar sub-branch, commonly known as MT (Machine Translation). Growing up in San Diego's suburbs, I always wanted to write, to travel the world beyond our Navy-bound shores, and to teach others about what I'd learned. During the latter half of the 1980s, as I wrote the concluding chapters of my doctoral dissertation—on the very un-modern topic of Restoration Drama—I came to know a young man, Michael ("Tac") Tacosky, who changed my life and the career choices I'd make for the next 15 years. This is the story about how I became involved in the fledgling MT business, from an operation I call "Tac's Garage" to an office of up to 35 people serving as a part of a division of several large, multinational corporations. This is the story of high hopes, naïve success, and a decline and fall brought about by the familiar demons of greed, shortsightedness and bad planning.

As an undergraduate at the University of San Diego in the mid-70s, I discovered I had a strong interest in languages, and double-majored in English and French. USD was a small, private Catholic university, and linguistics and languages were combined under a single department. USD's language programs were geared towards traditional "liberal arts" studies, not science. Therefore, my interests in a deeper understanding of the structure and function of language from a scientific viewpoint were never fulfilled.

Despite the factors above, I grew up in a family of early computer pioneers. My parents had worked at Naval Ocean Systems Center in Point Loma, and had been programmers since the late 1950s. Early memories still pop back into my mind of my mother bringing cartons of IBM punch cards home, to sort manually. Then there were the reams of line printer paper filled with obscure COBOL code, which we folded for Christmas decorations during the holidays....

In December 1986, I attended a Christmas party for the Catholic Young Adults group in a suburb of San Diego. Quite a few late 20s and early 30s computer professionals attended the party at the home of Joe Bowen, a young engineer. As I sat sipping my drink by the living room couch, a robot whirred nearby, terrorizing the Bowens' German shepherd puppy. On

the couch in front of me sat Tac, who was also a young engineer and contractor. Tac was a short, wiry young man of around 25. I had already known him for several years; we used to spend hours talking about politics, computers and languages, over chess games at my parents' house. He used to hold parties at his condo in Mira Mesa regularly, and really wanted play "the host" par excellence to our small social group. Nevertheless, it always seemed to me that Tac presented himself best when sharing his big ideas with smaller groups and individual friends. Tac was an activ-

ist, and he had traveled South and Central America, including the jungles of El Salvador. He embraced all the leftwing political causes of the time; what Reagan was for, he was against. He was sincerely full of nervous energy, so it was hard not to get caught in the grip of his enthusiasm.

While I sipped my New Coke on the rocks, I listened to Tac discuss the plight of the Sandinistas, and the dire situation in Mexico City after the earthquake. Then the conversation drifted to languages. Tac was a fluent non-native speaker of Spanish, and we started comparing Spanish to French.

"You know, Tac, French and Spanish have a lot in common. Take the verb conjugations, for example," I said.

"Actually, I've been thinking about that," he said. "Have you seen one of these before?" With that question, he slipped a paper contraption out of his pocket and handed it to me. It was a white cardboard wheel. On it were the major Spanish verb forms, with another wheel in the center, where the "user" could spin tense and verb forms to match the correct endings on the outer wheel.

I played with the wheel for a minute, lining up "yo" with "hablo" and "tu" with "hablas."

"You could do the same for French," I said. "You could line up 'je' with 'parle,' and 'tu' with 'parles,' and so forth."

"I've been thinking I could make a utility program out of this verb wheel," Tac said. He had an intense look in his deep-set eyes. "I want to make an electronic verb wheel."

This is the story of high hopes, naïve success, and a decline and fall....

Tac, at this time, was transitioning from working as a contractor with local computer firms, to run-

ning his own business. Tac and Joe Bowen, the engineer whose party provided some of the initial spark for MicroTac, had been working in Joe's garage on some engineering projects. My parents, through their business contacts, had gotten Tac and Joe a project called "Colorperm," where they were to create a calculator for hairdressers to choose die colors. After struggling for months to squeeze the necessary processing power

down to a puny (for the time) calculator-sized microchip, Tac and Joe were ready to give up on this project.

The Verb Conjugator

This verb conjugator became the first goal of what later came to be known as MicroTac Software. For the next year, I worked with Tac on his verb wheel project. He did the main programming of the initial program. The first Spanish Verb Conjugator was a TSR: Terminate and Stay Resident program. The idea was that a user could load the program within the 640K memory limit of DOS-based PCs in the 80s, and activate it with hot keys while simultaneously using a word processor. The users could highlight a word in their document, look up the correct form, and insert the conjugated form directly into their document.

The linguistic design of the program was straightforward: a three-dimensional array, with verb conjugation model codes along the top row, and the actual forms and endings corresponding to the models in rows and columns beneath. So, for Spanish, the verb “amar” would have a “Conj Code” of 01, with each tense of “amar” (present, preterit, imperfect, future, etc.) having a numbered column, which then referred to a numbered set of rows and columns along the bottom of the file. Amazingly, this old code, the primordial MicroTac code, survived all the ensuing years and changes that the company and its software underwent. It also remains, anecdotally, the single most popular feature of MicroTac’s programs, still sought out and used by translators in the 21st Century. My task for the project was to work with Tac in researching and inputting the data into the conjugation tables. There are standard reference works in most university bookstores, and we used more than one to create our models and populate the lexicon.

While Tac did most of the Spanish, I worked on the French system. He would bring his “luggable” 286 to my parents’ house, and I’d look over his shoulder as he did the initial programming. (“It fairly flies,” he used to tell me.) The programming language he used was Pascal. Later in 1988, since I had some experience in graduate school with German, I also added the initial data for the German verb forms. I was the first to tell Tac that my German wasn’t near the quality needed

for a commercial product, and I was not happy with the results. But this first attempt at the data became the predecessor of all our “troublesome” German products to follow. For design and philosophy reasons I’ll discuss in later sections, German was always our weak linguistic stepchild. As time passed, I also (in 1991) learned Italian, in order to help with the implementation of that system. Italian, too, was a problem area for a long time to come. As Tac did the initial programming, he would give me versions of the software to run on my own IBM XT. Because all the codes for the verb forms were in hexadecimal format, I had to run a calculator tool (yet another TSR) to determine which hex numbers to insert in the tables.

Researching the verbs and inputting the codes and corresponding inflections into the system was tedious and time-consuming. As 1988 merged into 1989, I found that helping my good friend with his little project was interfering with completing my PhD dissertation. Tac, however, was very enthusiastic about the potential of the eventual project, and offered me a “sweet deal” to continue my work. He offered me a contract for “one dollar for every French verb conjugator sold,” as well as enrollment in an Esperanto class taught in La Jolla, California. Although I never thought the project would make me rich, I signed the piece of paper Tac offered me; we each kept a copy. Both of us did take the Esperanto class, since Tac thought Esperanto linguistics might give us some insight into the languages we were working with. We never did anything with Esperanto other than study it, but the effort was sincerely naïve, like so much of our later work.

Tac tried two different methods of marketing the programs. First, he tried to sell them as reference tools to Houghton Mifflin’s textbook division (which had offices in San Diego at that time). When that deal fell through, he marketed the products as “shareware,” a popular formula for selling small software products in the 1980s and early 1990s. With shareware, the author of the software distributed it for free with a license, which obligated the user to purchase the product within a certain amount of time. To make

the shareware concept work for us, we needed to copy floppy disks: the more, the better. Just before working with Tac, I had been earning some extra money by helping my parents’ computer consulting business. One of the jobs my parents contracted for was copying floppy disks for several local companies, specifically a local printer manufacturer. For hours I would sit at the computer, textbooks at my side, copying 5.25” disks from drive A to drive B. So, when Tac began looking for a disk copier, I referred him to Golden Bow Systems, who manufactured the copying software my parents were using. Soon afterwards, Tac went to Golden Bow and met a person who would change the direction of his start-up company.

The childlike enthusiasm of the early days was about to get a ... dose of Business 101

This referral led to the next phase of MicroTac, the full development of Language Assistant as a

popular reference tool. A major factor in this development was when Gareth Juntunen from Golden Bow joined Tac as his partner and vice president. I never knew exactly how Tac lured Gareth away from Golden Bow. I only experienced how Gareth lent his marketing experience to Tac’s engineering talent, and the new MicroTac was born.

Physically, Gareth was a lot different from Tac. Where Tac was short, dark and shy, Gareth was tall, with red hair and an engaging smile. Gareth was outgoing, with none of the “nerdy” social awkwardness that seems to be a feature of many successful software entrepreneurs, from Bill Gates on down. Gareth had the business acumen to take Tac’s ideas and market them. He was ever the salesman, but also knew how to find the best in the products and get them to the public. One of the best early examples of this was Gareth’s reform of Tac’s business cards. The first MicroTac business card was a gaudy plastic 3D agglomeration of red, white and blue arrows and circles that shimmered as you moved the card under lights. After Gareth became a partner, he got Tac to dump the card in favor of a simple blue, black and white one. The childlike enthusiasm of the earliest days was about to get a serious dose of Business 101.



Translation Automation

...continued from page 9

cascaded supply chain. It is not uncommon that a localization project goes through three to four different levels of subcontracting.

The localization department in a customer organization hands out a 20-language localization project to the global multi-language vendor (MLV). The MLV analyzes the project, sets up the project plan and ships the project to smaller multi-language vendors or single-language vendors (SLV) who take care of linguistic testing and quality control. But quite likely they will contract out the actual translation work. Interestingly enough this cascaded supply chain has formed itself more or less around the phases of the evolution of our industry. The translators at the bottom of the supply chain are still doing the translation work. The larger regional and local localization vendors have taken on the tasks of engineering and managing the translation memory tools and the quality tracking. The global MLVs claim their value in this process as the process experts and high level project managers. Quite naturally every link in this chain costs time and money. A closer examination of the activities in the entire supply chain shows that there is a tremendous overlap in tasks between the different actors. The customer will perform some level of quality control, project management, translation memory management, and so will the global MLV. The regional MLV and the local SLV will also perform quality control and project management tasks and they will also maintain translation memory databases for the project. And finally the translators will of course control the quality of their work and work with translation memory tools.

History of Translation Automation

The idea of automating translation has fascinated people for a long time. In 1954 the first experiment to have a computer fully automatically translate a text from Russian into English was called a big success. And within a few years research teams were put to work in multiple places around the world to let computers crack this big problem. The initial euphoria turned into great disillusionment when in

1966 the ALPAC report concluded that fully automatic high quality translation was not likely to happen in the foreseeable future. However government agencies continued to test and use machine translation for content scanning and translation support. At the start of the localization phase in the mid eighties the debate about machine translation for commercial use heated up. Customers and translators reviewed the few commercial machine translation products that were available at the time. The sudden big volumes of translation work urged some level of automation of the translation process. But as the quality standards were very high at that time the option to use machine translation was rejected. The idea was born to let computers help the human translators by storing translated segments and terms in a database and offering them for reuse when applicable. The first so-called translation memory tools came on the market in the early eighties. They were sold as productivity tools for individual translators. Some of the biggest clients of the localization industry adopted the TRADOS Translators Workbench and imposed it on their vendors. Within a few years TRADOS became the de facto standard for the localization industry. Although today well over 30,000 of these packages have been sold, translation memory software has also displayed its severe limitations to the enterprise customers. In a cascaded supply chain with sometimes over one hundred translators working on a project into one language for one customer it is very inefficient to work with desktop productivity tools. Every contractor and the ultimate client maintain copies of the translation database. Discrepancies are inevitable, even though special dedicated functions are created for cleaning, copying and updating translation memory files at every hand-off in the process.

Shorter updating cycles and the growing complexity of localization work made it necessary to centralize and automate the process. It became apparent that the cost of managing the process exceeded the cost of the actual translation. New software products were introduced in the localization industry in the late nineties that would manage the entire workflow and integrate translation memory functionality. This software was called globalization

management software (GMS). The first GMS products appeared to be too rigid. They would force customers to work in a strict predefined process and often effectively locked them in with a translation supply base with the service vendor selling the GMS software. However, these products provided an answer to the problem of project and file management. And with this, the point of pain came back to the increasing volume and shortening timeframes for the translation itself.

The GIL Market

According to an IDC survey of the globalization, internationalization and localization (GIL) market the industry is growing from \$4.2 billion in 2001 to \$8.9 billion in 2006, representing an annual growth rate of 16.3%. Localization and translation services form by far the largest part of this market with 69.8% of the total or \$2.9 billion in 2001 and growing to \$5.8 billion in 2006, representing an annual growth rate of 14.6%. IDC also distinguishes a sub-sector for globalization strategy and internationalization consulting, which they expect to grow faster than the localization and translation services sector, reaching a total of \$2.9 billion in 2006, or 39% of the total. The fastest growing sub-sector will be cross-lingual applications. Representing less than 1% of the total market in 2001 or \$42 million, IDC expects this sub-sector to grow to \$193 million by 2006, or 35% annual growth.

Big Changes

Fifty years of history in the translation industry and fifty years of research and development in translation automation have generated very few results. The key inhibitor for translation automation has been the lack of a quality standard. People are emotional about language and instinctively go for the maximum quality level without being able to measure it. But that is about to change.

Until recently, localization was hardly a strategic issue for corporations. For most companies it was just seen as the cost of doing business overseas. The ROI calculation was easy to make: the additional revenue of a localized product must exceed the cost of localization. Corporate decision makers are now realizing that the

Continued on next page ►

EAMT/CLAW

...continued from page 2

quality of the reference translations in the system database; the more these are controlled, the better the expected quality of translation output by the system.

Conference Goals

The aim of this conference is to provide a forum in which the problems accompanying controlled language translation may be outlined, possible solutions proposed, and in general to bring together developers, implementers, researchers and end-users from the publications, authoring, translation and localization fields to discuss how ideas from both the authoring and translation camps might be integrated in this common area.

Categories for Submission

Controlled translation: What is controlled translation; RBMT and controlled translation; TM/EBMT and controlled translation; Influence and interplay of controlled language upon both source-language parsing and target-language generation in an MT system; Role of the lexicon in controlled translation; Can we expect better controlled translations from a hybrid approach? Or from a multi-engine approach? Towards a Roadmap for controlled translation - the way ahead?

Machine Translation: MT for the Web; Practical MT systems; Methodologies for MT; Speech and dialogue translation; Text and speech corpora for MT and knowledge extraction from corpora; MT evaluation techniques and evaluation results; MT postediting.

Controlled Language: Examples of controlled languages: their definition, by whom, and intended usage; Consequences for technical authors and implications for Natural Language Processing; Practical experiences of teaching and using controlled languages; Application of controlled languages in speech systems.

System Demonstration: Abstracts for demos must not exceed 400 words. Developers should outline the design of their system and provide sufficient details to allow the evaluation of its validity, quality, and relevance to controlled language. Pointers to web sites running the demo preview and/or screen camcorder video clips will also be helpful.

Invited Speakers

We are pleased to announce that invited speakers for the conference will include Steven Krauwer, University of Utrecht and Coordinator of ELSNET, and Lou Cremers, Océ Technologies.

The conference co-chairs are John Hutchins (WJHutchins@compuserve.com), on behalf of the EAMT, and Arendse Bernth (arendse@us.ibm.com), on behalf of CLAW. For details see www.eamt.org/eamt-claw03/. □

InstantService

...continued from page 3

nical support available using internet-only channels is important where customers are using dialup access to the internet. Web-based interaction allows them to stay online while asking questions and getting support. This is most important for consumer products, and regions of the world where broadband internet access is not common.

InstantService's business case is based on the estimate that the worldwide CRM (Customer Relationship Management) market was worth US\$ 25 Billion in 2002 (Gartner Dataquest). CRM typically includes sales support database tools, as well as contact centers. For comparison, the worldwide translation market has been estimated at \$11 Billion (ABI). Translation solutions offered by Instant Service are provided by WorldLingo for French, German, Italian, Brazilian Portuguese, and Spanish.

See: www.instant-service.com. □

Translation Automation

...continued from previous page

translation of product documentation is far from enough to compete in the global market. They are facing the challenge of globalizing their companies in their entirety. This will lead to changes in the GIL market that will overwhelm the capabilities and expectations of the current suppliers, large and small. First of all customers will be looking for hybrid solutions of human and machine translation. Depending on the required quality level, texts will be routed to a human translation supplier or to a machine translation engine. Customers will adopt simple quality metrics (like the SAE J2450 standard) to measure the quality of translations. Early results have shown that although customers' first impulse is to demand the highest possible quality, when confronted with an objective way of measuring the cost of perfection, most clients discover the measured level of quality that meets their customers' requirements, and their translation budgets. The acceptance criteria will be expressed in an exact scoring rate according to this metric. At the same time, deployment of

machine translation engines will lead to rapid improvement of the quality of output in a customer-specific environment, as all not-found words will be coded and added to the machine translation dictionaries. Customers in vertical industries will meet and start discussing unified terminologies, which will further enable the development of high-quality industry-specific translation systems. The Localization Industry Standards Association (LISA) was founded in 1990 to do exactly that: standardize computer terminology in multiple languages. Although it appeared to be too early days at that time, now corporations recognize the value of standardization. Unified terminology combined with the XLIFF standard (XML Localization Interchange File format) and the upcoming Translation Web Services (TWS) standard will lead to a much more transparent industry. As a result we will see open marketplaces where customers can search for the fastest, most economic supply of translation that meets the standard quality scoring rate they require.

Continued on page 18 ►

Conferences and Events

MT Summit IX

**New Orleans, Louisiana, USA
September 23-27, 2003**

The ninth IAMT Summit is taking shape! If you receive the electronic version of this newsletter, you still have time to submit a research paper, user study, or system description for inclusion in the conference program. The submission deadline is May 11, 2003. Please see details at: www.mt-summit.org.

Invited Speakers

MT Summit IX will feature three keynote presentations from the world's experts in the areas of research, development, and use of translation technology:

- Pierre Isabelle, Area Manager, Content Analysis, Xerox Research Centre Europe, Grenoble: "Multilingual Document Processing at XRCE";
- Akitoshi Okumura, Senior Manager, Human Language Technology Group, NEC Corporation, Japan: "Development of Speech Translation for Hand-held Devices";
- Donald Barabé, Director, Business Development, Translation Bureau, Public Works and Government Services Canada: "Soaring Demand, Shrinking Supply in Translation: How We Plan to Make Ends Meet"

Proposals for Panels and Special Sessions

The program committee invites members of the MT community to suggest panel topics (particularly if you can also suggest some qualified panelists) or special session topics to explore issues that are timely and may benefit from more in-depth coverage in the context of the conference. Submit proposals to the program chair, Elliott Macklovich, mtix@IRO.umontreal.ca.

Workshops

Four full-day workshops will allow participants to explore topical subjects in greater depth:

September 23, Pre-conference workshop

- MT Evaluation

September 27, Post-conference workshops:

- Semitic Languages;
- Teaching MT and Computer-Aided Translation;
- Interlinguas.

For more information, including contacts for each workshop's organizers, contact Harold Somers, workshops chair, harold.somers@umist.ac.uk.

Tutorials Proposals

Pre-conference tutorials are half-day or full-day intensive sessions led by experts from academia and industry. Tutorials offer the opportunity to catch up on the latest trends and ideas from research and industry, to get practical ideas

from expert users, or just get the kind of general

background on MT that will make the rest of the conference more understandable! Proposals for tutorials should be addressed to tutorial co-chairs: Laurie Gerber, gerbl@pacbell.net and Keith Miller, keith@mitre.org.

Exhibits

The MT Summit is a particularly good opportunity to get broad exposure to potential customers from industry and government from all over the world. In addition to MT developers and vendors, we are particularly eager to include: consultants and integrators who can help potential users to understand and make the best use of the technology that is available; providers of companion technologies that can leverage or provide leverage to machine translation, such as content management, globalization management systems, translation memory,

and others; Providers of component technologies that may be of interest to other developers and researchers, such as morphological analyzers and generators, parsers, and others. The conference exhibition will run from Tuesday, September 23 (opening at 6:00 p.m. during the conference reception) to Friday, September 26 (closing at 5:30 pm when the conference concludes.) Contact Information

Please contact: Laurie Gerber, gerbl@pacbell.net or Keith Miller, keith@mitre.org, exhibition coordinators, and see the conference website for details and updates.

The Conference Venue

The conference venue is the elegant Fairmont New Orleans hotel. It offers a stunning environment for a conference and is within walking distance of the famous French Quarter. Other accommodation is within two blocks. See www.fairmontneworleans.com.

About New Orleans

The city was founded in 1718, has a current population of 496,000, and is five feet below sea level. It's well known for its night life (bars are open 24 hours a day), European flair, and jazz. Tours are available through the local swamps, plantation houses, cemeteries and "haunted places." Available city tours allow the visitor to take in the French Quarter, Mississippi River, cemeteries and streetcars, all in a single day.

Many public events occur throughout the year, and several will take place around the time of the MT Summit. For more information, see www.neworleans.com.



Summit IX Important Dates

Submission deadline	May 11, 2003
Notification to authors	June 30, 2003
Final papers due	July 31, 2003

ACL 2003

Sapporo, Japan
July 2-12, 2003

376 papers have been accepted for presentation at this year's ACL conference. In addition to regular paper presentation, the conference will include poster and demo sessions, product exhibition, and a student research session, a number of pre-conference tutorials, and post-conference workshops, as well as related events nearby. The recently announced tutorials, workshops and related events are outlined below.

Tutorials

Four pre-conference tutorials will take place on July 7.

Finite State Language Processing: Gertjan van Noord (University of Groningen, the Netherlands)

Maximum Entropy Models, Conditional Estimation, and Optimization without the Magic: Dan Klein and Christopher D. Manning (Stanford University, U.S.A.)

Knowledge Discovery from Text: Dan Moldovan (University of Texas at Dallas, U.S.A.) and Roxana Girju (Baylor University, U.S.A.)

Spoken Language Processing: Separating Science Fact from Science Fiction: Roger K. Moore (20/20 Speech Ltd, U.K.)

Workshops and Related Conferences

Following the main conference will be 10 Workshops. Some of the workshops are accepting submissions into April. A selection are mentioned here, together with the workshop chair/co-chair names. *See the ACL website for a complete listing, workshop descriptions and submission details: www.ec-inc.co.jp/ACL2003/.*

July 7, 2003

The Sixth International Workshop on Information Retrieval with Asian Languages (IRAL2003). Tetsuya Ishikawa (Univ. of Tsukuba, Japan), Jun Adachi (NII, Japan), Kam-Fai Wong (The Chinese Univ. of Hong

Kong, China) Submission deadline: April 15. *See: research.nii.ac.jp/IRAL2003/.*

July 11-12, 2003

The Eighth Conference on Empirical Methods in Natural Language Processing (EMNLP2003), Michael Collins (MIT AI Lab, U.S.A.), Mark Steedman (University of Edinburgh, Scotland). Submission deadline: April 4. *See: www.ai.mit.edu/people/mcollins/emnlp03.html.*

Multilingual Summarization and Question Answering - Machine Learning and Beyond, Abraham Ittycheriah (IBM T.J. Watson Research Center, U.S.A.), Tsuneaki Kato (The University of Tokyo, Japan), Chin-Yew Lin (USC/ISI, U.S.A.), Yutaka Sasaki (NTT Communication Science Laboratories, Japan), Submission deadline: Apr 21, 2003

Second SIGHAN Workshop on Chinese Language Processing, Qing Ma (Communication Research Lab, Japan), Fei Xia (IBM, U.S.A.), Submission deadline: March 10, 2003

Towards a Resources Information Infrastructure, Steven Krauwer (ELSNET, the Netherlands), Nicoletta Calzolari (ILC-CNR, Italy), Antonio Zampolli (Università di Pisa & ILC-CNR, Italy), Submission deadline: April 13, 2003

July 11, 2003

The Lexicon and Figurative Language, Alan Wallington (University of Birmingham, U.K.), Submission deadline: April 13, 2003

The Second International Workshop on Paraphrasing: Paraphrase Acquisition and Applications, Kentaro Inui (NAIST, Japan), Ulf Hermjakob (USC/ISI, U.S.A.) Submission deadline: April 21, 2003

July 12, 2003

Multilingual and Mixed-language Named Entity Recognition: Combining Statistical and Symbolic Models, Mari Broman Olsen (Microsoft Corporation, U.S.A.), Submission deadline: April 4, 2003. □

Aslib Translating and the Computer 25 - Conference and Exhibition

London, UK
November 20-21, 2003

Call for Papers

This conference is one of the few international events which focuses on the user aspects of translation software and as such has been particularly beneficial to a very wide audience including translators, business managers, researchers and language experts.

Once again, this year the conference will address the latest developments in translation (and translation-related) software. It will address the needs of the following conference attendees: industry, public administration, agencies, freelancers, development

This call for papers invites abstracts of papers to be presented at the conference. The papers (and the presentations) should focus on the user aspects of translation or translation-related software rather than on theoretical issues. Presentations accompanied by demonstrations are especially welcome.

Topics

The range of topics includes (but is not limited to): use of MT systems, machine-aided translation and translation aids, controlled languages and their use in MT, speech translation, terminology, localization, multilingual document management/workflow, case studies of technology-based solutions, the Internet and translation aids/services, the value of "free" versus "charging" services/sites on the Internet.

Submission Guidelines

Authors are required to submit an abstract of a minimum of 500 words of the paper they would like to present, together with an outline of the structure of the paper and short biography. Abstracts should be sent by post or email to:

Continued on next page 17 ►

The LISA Forum Europe

June 30 - July 3, 2003

Radisson Edwardian Heathrow
London, UK

Global Content Management is today's number one international business challenge. So you've entered new markets and are operating worldwide, but are you satisfied with your ROI? Successful global companies (from the established automotive and banking industry to high tech software) are looking for ways to increase their company's production for faster delivery and turn around time. Global workflow is a fundamental strategy in achieving this.

The 2003 LISA Forum Europe international conference will focus on how companies are leveraging cultural, product, customer, and financial knowledge gained through existing operations to explore new business and facilitate market expansion. The program will feature:

- Globalization Resource Planning: Preparing for the Future
- System Integrator Alliance Strategies: Developing an Effective Network
- Cutting Translation and Localization Costs
- Greater Customer Satisfaction Through Quality Improvement Metrics
- Tools Benchmarking by Comparing Best Technologies
- Implementing Knowledge for Worldwide Training: GMS = A Fast Track Education
- Global Content Management - What is being delivered today and how!
- System Integration - Content Creation, Translation and Localization
- Vendor Selection - Choosing the Partner and Technology that is Right for You!

- Globalization Consultancy - Managing Implementations Successfully
- Leveraging Global Customer Support to Expanding Market Presence and Revenues

Who Should Attend

C-Level Executives, International Managers, Enterprise Globalization Strategists, Company Directors, Knowledge Managers, Global Web-Development Specialists, Project Managers, and Localization Production Managers from both client and vendor companies seeking to learn more about how to implement language processing standards effectively, efficiently while lowering expenses and providing revenue opportunities

LISA Forum Europe attendees are decision makers and key suppliers, responsible for global information management systems, or the recommendation of the localization and translation operations for software publishing companies, telecom and Internet-service providers, localization service companies and internationalization consulting service groups.

Automated Workflow Tools and Localization Service Exhibition

The industry's leading technology developers for automated language processing systems, multilingual web-services, internationalization, translation and localization suppliers will demonstrate their products and services.

The result is 4 days of dedicated focus on the optimization of language processing standards, with feedback from the business professionals who have done it and the technology providers who can design the right system for your company – from content creation through to web-services solutions. See: <http://www.lisa.org/events/2003uk/>. □

LISA Important Dates

Early Bird Registration May 15, 2003
Registration Deadline June 25, 2003

LangTech 2003

November 24-25, 2003

Meridien Montparnasse Hotel
Paris, France

LangTech 2003, the second edition of the European forum for language technology, offers a unique platform for the Language and Speech technology community - developers, integrators, entrepreneurs, adopters and users.

The conference topics address the latest technology breakthroughs and are presented by experienced practitioners in the field - either adopters/users stressing the feasibility of the technology, or by solution providers presenting successful case studies.

Practical solutions and near-market research results are presented during the conference and in the accompanying exhibition, showing state-of-the-art products and solutions

LangTech 2003 will feature keynote talks from leading players, presentations from a wide range of developers and solution providers, panel discussions of key issues affecting the market in Europe and beyond, and an exhibition of applications, products, services and research prototypes.

Special sessions will enable start-up companies to promote and pitch their products and services and explore funding possibilities. LangTech 2003 will also offer pre-conference tutorials on new methods and hot technology developments. Demonstrations of applications, products, services and research prototypes will be featured in the exhibition.

LangTech 2003 is organised by ELDA, the Evaluations and Language resources Distribution Agency, assisted by the Euromap committee and several other corporate sponsors.

For program, registration and venue information, visit the LangTech 2003 Web site at: <http://www.lang-tech.org/default.html>. For ELDA information, see <http://www.elra.info/>. □

Building and Using Parallel Texts: Data Driven MT and Beyond

An HLT-NAACL 2003 Workshop
Edmonton, Alberta, Canada
May 31, 2003

The goal of this workshop is to provide a forum for researchers working on problems related to the creation and use of parallel text. Recent events have demonstrated once again the importance of inter-language communication, and reinforce the need for advances in machine translation (MT) and multi-lingual processing tools.

The workshop will be centered around the problem of building and using parallel corpora, which are vital resources for efficiently deriving multi-lingual text processing tools. In addition to regular papers, the workshop also includes a shared task that will result in a comparative evaluation of word alignment techniques. In addition to research papers, the workshop will include an invited presentation by Elliot Macklovitch, of RALI, University of Mont-

ASLIB

...continued from page 15

Nicole Adamides, Conference Organiser; Aslib, The Association for Information Management; Temple Chambers, 3-7 Temple Avenue, London, EC4Y 0HP
Tel: +44(0) 20 7583 8900; Fax: +44 (0) 20 7583 8401; Email: nadmides@aslib.com
Programme Chairs: Daniel Grasmick and Chris Pyne, SAP; Professor Ruslan Mitkov, University of Wolverhampton and Olaf-Michael Stefanov, United Nations.

ASLIB Important Dates

Submission deadline	June 15, 2003
Notification to authors	August 1, 2003
Final papers	October 10, 2003

real, and the results of an evaluated shared task.

Submissions were solicited on the following issues:

Construction of parallel corpora, including the automatic identification and harvesting of parallel corpora from the Web; Methods to evaluate the quality of parallel corpora and word alignments; Tools for processing parallel corpora, including automatic sentence alignment, word alignment, phrase alignment, detection of omissions and gaps in translations, and others; Using parallel corpora for data driven Machine Translation; Using parallel corpora for the derivation of language processing tools in new languages; Using parallel corpora for automatic corpora annotation; Language learning applied to parallel corpora; Translation memory systems as a source of aligned corpora

Shared Task

Researchers who have word alignment systems available were invited to participate in a shared task, individually or as part of a team. Participants in the shared task were provided with common sets of training data, consisting of Romanian-English and French-English parallel texts. Following a one-month training period, participants ran their alignment systems on test data provided, submitting their results for evaluation using a common set of metrics. The results will be reported at the workshop.

Workshop Co-chairs: Rada Mihalcea (University of North Texas), Ted Pedersen (University of Minnesota, Duluth)
See: www.cs.unt.edu/~rada/wpt.

JHU Summer Workshop: Syntax for Statistical MT Baltimore, Maryland, USA

From roughly 1988-1994, the DARPA HLT program sponsored seminal NLP research that yielded the IBM statistical MT models, and prototype translation systems, among other things. While component techniques quickly were adopted and became influential in many areas of NLP re-

search, full statistical MT seemed to go dormant until the 1999 Johns Hopkins University summer workshop in which the world's top researchers devoted 6 weeks to reimplementation of the IBM translation systems, and development of toolkits to foster additional work in this area.

The JHU summer workshops have continued, touching on a variety of topics. This year's MT workshop will bring together 8 international researchers, and 8 undergraduates now being recruited in a nationwide search, to work on Syntax for Statistical Machine Translation. Three other workshops conducted in parallel will study: Semantic Analysis Over Sparse Data, Dialectal Chinese Speech Recognition and Confidence Estimation for Natural Language Applications. Note below that many of the opening and closing overview presentations are open to the public.

Workshop Dates

The workshops will be conducted June 30 to August 22, 2003 at Johns Hopkins University, Baltimore, Maryland, USA. Mixed teams of leading professionals and students will fully cooperate to advance the state of the art. The professionals are university professors and industrial and governmental researchers presently working in widely dispersed locations. Eight undergraduates have been selected through a nationwide search from the current junior class based on outstanding academic promise. Graduate students familiar with the field have been selected in accordance with their demonstrated performance.

Programs

The Opening Day Program begins at 9:00 AM on June 30, 2003 in Arellano Theater on the JHU Homewood Campus. The program is open to the public. A detailed agenda for WS03 Opening Day will soon be available. Speakers on Opening Day will include government sponsors and personnel, team leaders, and technical and administrative personnel. The program usually lasts until 12:00 noon.

The Closing Day Program will be held over two days August 21 and 22, 2003. The program is open to the public. A

Continued on page 19 ►

Data-Driven MT Grows Up

LG

Following last year's AMTA conference (October, 2002, Tiburon, California), it became clear that a number of data-driven MT companies were on the verge of commercializing new machine translation systems. In parallel with this, established developers in the U.S. Japan, and Taiwan were starting over with new hybrid engines slated to replace their earlier symbolic engines. Early this year, we began contacting the upstarts to find out what we could expect from them and when. We asked each developer questions, including "what is your core approach to machine translation" and "what sets it apart from other machine translation systems". Two surprising pieces of information emerged:

1: There are quite a number of upstarts! We found 6 startups with the primary mission of commercializing data-driven MT systems (Aixplain, AKS, Huajian, Language Weaver, Linear-B, Verbalis). A number of other companies with broader offerings (or projected offerings) have entered the ring with data-driven MT systems (Meaningful Machines, Microsoft, Morphologic, Symbionautics). And of course a number of the existing MT developers are actively introducing data-driven methods into existing rule-based products (Systran), or working on completely new products with more of a data-driven core (Bowne Global Solutions (the Barcelona system), and BehaviorTrans in Taiwan).

2: The designs of the systems challenge the distinction established in 1992, between the empiricist and rationalist approaches to MT. The rationalists, we understood, built systems with hand-crafted rules. The empiricists used automated learning techniques to extract information from linguistic data, primarily existing translations. But in the current group, there are example-based systems where the examples are all hand-crafted from existing translations (Verbalis, Oki, MorphoLogic, IBM Japan). The research community may object to the use of the term "example-based" in this case, but we

have used the descriptions provided by developers. In addition, many of the systems claim to be hybrids of various sorts (Aixplain, Microsoft, Oki, Sehda, Huajian and Linear-B).

We have not attempted to evaluate developers' claims regarding the type of system being developed. Although the MT research community seems to have a clear understanding of the necessary and sufficient conditions for considering a system "statistical" or "example-based," those outside the ACL/AAAI research community have adopted and applied the labels more loosely.

Because of space limitations, the results of the survey will be presented as a two or three part series. Readers are invited to

suggest systems that should be covered in this series.

Company: Aixplain

Aixplain's core approach to MT is Statistical and symbolic MT (a complex hybrid system). Its target market is Business clients, with a "B-2-B" approach.

What sets your system apart from other MT systems?

- very high flexibility
- very low Time-To-Market
- high quality of translations within specific domains
- adaptive HLT system

Aixplain AG
Monheimsallee 22
52074 Aachen, Germany
Tel: +49.241.18927-0
www.aixplain.de

Data-Driven MT Companies at a Glance

Company: Aixplain

Founded: 2001
Inventor: Hassan Sawaf, Hermann Ney and Franz Josef Och
CEO/President: Hassan Sawaf (CEO), h.sawaf@aixplain.de
Customer/Investor contact: Chafik Moalem, c.moalem@aixplain.de
Company size: 22 people
First Product deployment: October 2001

Company: IBM Japan

When did IBM Japan start developing empirically-based MT systems? 1990
When did IBM Japan start its NLP R&D group?

- Early 80s (MT research), 1970s (general NLP research)

Who is the inventor of the empirical technology that has been commercialized?

- Koichi Takeda

Leader of the MT development group/NLP R&D Group at IBM Japan:

- Dr. Hideo Watanabe, hiwat@jp.ibm.com

Sales Information (in Japanese): www-6.ibm.com/jp/software/internet/king/
First sale or deployment: "Internet King of Translation" released by IBM Japan in 1996.

Company: Fluent Machines

Founded: Fluent Machines, a subsidiary of Meaningful Machines, was founded in 2001.
Inventor: Eli Abir, Inventor and Chief Architect
CEO and Chairman: Steve Klein
President: David Miller
Customer/Investor contact: Michael Steinbaum, COO, mike@meaningfulmachines.com
Company size: 11 employees
First product sale/deployment: fourth quarter of 2003 (projected)

Company: Linear-B

Founded: 2002
Inventors: Colin Bannard and Chris Callison-Burch
Customer/Investor Contact: Colin Bannard, colin@linearb.co.uk
First product deployment/sale: end of 2003 (projected)

More DDMT systems next issue!

Company: IBM Japan

IBM Japan's core approach to machine translation is Pattern-based. Usually empirical approaches include example-based approach and statistical approach. In this sense, IBM has not yet released any empirical-based commercial MT system. They report their target market or application to be Web browsing.

But, in a broad sense, the pattern-based approach which IBM Japan developed can be considered as one of the empirical approaches. IBM Japan's "pattern-based approach" uses a large number of translation patterns, each of which is a pair of source lexicalized CFG rule and a target lexicalized CFG rule. It, given a source sentence, analyzes it by using CFG parsing method with source side rules of translation patterns, and generates a target structure by synchronized derivation mechanism using translation patterns whose source rules are used for parsing. We collected about more than 10,000 translation patterns. For technical details, please see the following paper: Takeda, K., "Pattern-Based Context-Free Grammars for Machine Translation," Proc. of 34th ACL, pp. 144--151, June 1996.

What sets your system apart from other MT systems? The pattern-based approach allows users to add phrasal-level translation knowledge as well as word-level knowledge. For instance, a user can register a Japanese translation for a phrase "hit a big shot."

Dr. Hideo Watanabe
Group Leader of Intelligent Information Human Interaction Technology, S&S,
Tokyo Research Laboratory
IBM Japan
Tel: +81-46-215-4561

Company: Fluent Machines

The Fluent Machines approach to machine translation incorporates several purely empirical natural language learning processes that are new to the fields of machine translation and NLP.

Fluent Machines expects their first applications to target major European languages, followed by major Asian languages.

What sets your system apart from other MT systems? Eli Abir, inventor and architect of Fluent Machines' technology, had important insights into the way different

languages represent the same universal ideas. Mr. Abir leveraged those insights into four novel and unique processes: (i) the first process uses previously translated parallel text to automatically build large cross-language databases of basic word-string combinations, (ii) the second process leverages known word-string translations between language pairs to discern translations between different language pairs, (iii) the third process determines semantic equivalents in both the source and target languages, and (iv) the fourth process links together target language word-string translations created by any of the other processes to produce translated text.

Furthermore, Fluent Machines leverages a patent-pending technology that truly understands wide-ranging natural language through a process that (1) uncovers language patterns found in written text (any language and across all domains) and (2) focuses on concepts regardless of the words used to express them. This enables the Fluent Machines system to understand single and multi-word concepts, which gives tremendous power and flexibility to the MT system.

1450 Broadway, 40th Floor
New York, New York 10018
Tel: 212-716-0070
info@fluentmachines.com

Company: Linear-B

Linear B's core approach to MT is a combination of machine-learning and example-based methods. Their target market/application is robust translation of web content and email.

What sets this system apart from other MT systems? Our system has been designed from the outset to deal with, and indeed utilize, the diverse nature of web data and the specific needs of web users in a more satisfactory fashion than the one-size-fits-all approach of most systems. It employs innovative techniques for dealing with the non-stationary nature of language, and crucially for acquiring knowledge from previously neglected kinds of data, meaning that it is able to offer the advantages of robustness and low cost that are associated with data-driven systems for a larger number of language pairs than have previously been possible.

Linear B Ltd.
Edinburgh Technology Transfer Centre
King's Buildings
Edinburgh EH9 3JL, Scotland
Tel: +44 131 472 4816
Linearb.co.uk

Continued in next issue... □

Translation Automation

...continued from page 13

About the Author

Jaap van der Meer is co-founder of Cross Language, a company that specializes in the implementation of machine translation and other cross-lingual applications. Before that he was CEO of ALPNET from 1995 until 2001. In 1980 Jaap van der Meer started a pioneering localization company, INK International, in The Netherlands. In the late eighties and early nineties he published Language Technology, an innovative magazine covering the early developments in speech, language and translation technology. In 1990 he co-founded LISA, the Localization Industry Standards Association. In 1999 he co-founded the TopTec conference for multilingual communications in the automotive industry. □

JHU Summer Workshop

...continued from page 17

detailed agenda for WS03 Closing Day Program should be available by the beginning of August. The Closing Day Program allows the teams to present their progress and results to the language and speech community. Each day's program runs from 9:00 AM to about 4:30 PM.

Guest Lecturers

There will be several guest lecturers over the workshop period. Guest Lectures are scheduled at Arellano Theater on the JHU Homewood Campus. Details on the speakers, titles, and abstracts will be posted as information becomes available.

See: www.cisp.jhu.edu/ws2003/. □



MT News International

Subscription Order Form for non-members of IAMT Associations

Subscription to **MTNI** is a benefit of membership in any of the three regional IAMT Associations. Non-members may also subscribe. This form should be sent to the appropriate region together with a remittance in the currency specified. The fee covers a one-year airmail subscription (individual or institutional) for three issues, starting in the spring of the current year.

I/we wish to receive a one-year subscription to MT News International:

Name: _____

Organization: _____

Address: _____

City: _____ State: _____ Postal code: _____

E-mail: _____ Fax: _____

For individuals or institutions located in the Asia-Pacific region, please return this form with payment of ¥4,000 (bank draft or international money order) to:

Association for Machine Translation in the Americas
PMB 300
1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

For individuals or institutions located in Europe, the Middle East, or Africa, please return this form with payment of Sw.fr. 70.00 (check or money order) to:

Association for Machine Translation in the Americas (AMTA)
PMB 300
1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

For individuals or institutions located in North, South, or Central America, please return this form with payment of US\$ 75.00 (check, money order, or credit card) to:

Association for Machine Translation in the Americas (AMTA)
PMB 300
1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

Publications Order Form

Please return this form with payment or credit card information, to:
International Association for Machine Translation (IAMT c/o AMTA)
PMB 300, 1201 Pennsylvania Avenue, N.W. , Suite 300
Washington, DC 2004 USA

Please send the items marked at the right to:

Name: _____

Organization: _____

Address: _____

City: _____ State: _____ Postal code: _____

E-mail: _____ Fax: _____

Price (in U.S. dollars)¹

Title	Member ²	Non-member
<input type="checkbox"/> Compendium of Translation Software (on-line version)	FREE	\$20.00
<input type="checkbox"/> Proceedings of MT Summit VI	\$40.00	\$60.00
<input type="checkbox"/> Proceedings of AMTA-96	\$40.00	\$60.00
<input type="checkbox"/> Proceedings of AMTA-94	\$40.00	\$60.00
<input type="checkbox"/> Proceedings of Workshop on MT Evaluation (1992)	\$55.00	\$55.00

¹ Prices include shipping and handling.

² Member of AAMT, AMTA, EAMT.

Method of Payment

Check or M.O. Visa MasterCard American Express

Card number _____ Exp. Date: ____/____

The proceedings of AMTA-98 and AMTA-2000 appeared as #1529 and #1934 in the Springer series Lecture Notes in Artificial Intelligence. To order, contact the publisher at www.springer.de.

Asia-Pacific Association for Machine Translation

CORPORATE / INDIVIDUAL APPLICATION FORM

Please fill out the appropriate form in both your native language (NAT) and English (ENG) and send it to the address below.

CORPORATE

Corporate name: _____ Seal: _____
 NAT: _____
 ENG: _____
 Capital: _____ Date established : _____ # employees: _____
 Name of company president: _____ Seal: _____
 NAT: _____
 ENG: _____
 Name of person responsible for this application: Seal: _____
 NAT: _____
 ENG: _____
 His/her office/department: _____
 NAT: _____
 ENG: _____
 Mailing address: _____
 NAT: _____

 ENG: _____

 Tel: _____ Ext.: _____ Fax: _____
 E-mail: _____
 Business category:
 Government agency Manufacturer
 Service industry Translation business
 Other (please specify): _____
 All corporate application forms should include a company prospectus.

INDIVIDUAL

Name: _____ Seal: _____
 NAT: _____
 ENG: _____
 Mailing address: _____
 NAT: _____

 ENG: _____

 Tel: _____ Fax: _____
 E-mail: _____
 Occupation: _____
 Company name: _____
 Company address: _____
 Translator Other: _____
 University/Institution/Researcher
 Specialty: _____

Individual Membership Fees

Initiation fee for individual members	¥ 1,000
Annual dues for individual members	¥ 5,000
Total payment	¥ 6,000

The foregoing amount will be paid by ___/___/___, or within one month from the date this form is mailed.

Corporate Membership Fees

Initiation fee for corporate members
 (1 unit = ¥10,000; minimum 1 unit) ___ units = ¥ _____
 Annual dues for corporate members
 (1 unit = ¥50,000) ___ units = ¥ _____
 Please check appropriate box below:
 MT system developer (minimum 10 units)
 Other, capital over ¥10 million (minimum 2 units)
 Other, capital up to ¥10 million (minimum 1 unit)
Total payment (initiation fee + annual dues) ¥ _____
 The foregoing amount will be paid by ___/___/___, or within one month from the date this form is mailed.

Method of Payment

Wire transfer to:
 Mizuho Bank
 Ochanomizu Branch
 Tokyo, Japan
 A/C No. 1737479
 For: Asia-Pacific Association for
 Machine Translation (AAMT)
 International postal money order in
 Japanese yen to AAMT at ▶
 Applicants are responsible for all bank
 charges. Please retain your copy of bank
 draft or money order as proof of payment.

Today's date: ___/___/___

Please send this form to:

Asia-Pacific Association for Machine
 Translation (AAMT)
 c/o Japan Electronics and Information
 Technology Industries Association (JEITA)
 Mitsui Kaijo Bekkan Building, 3F
 3-11, Kanda-Surugadai, Chiyoda-ku
 Tokyo 101-0062, Japan
 Fax: +81 (03) 3518-6472

Association for Machine Translation in the Americas

MEMBERSHIP APPLICATION / RENEWAL FORM

Type of member and membership fee per calendar year:

- Individual US\$ 60
 Institutional (nonprofit) US\$ 200
Representative: _____
 Corporate US\$ 400
Representative: _____

Please return this form, together with your payment or credit card information, to:

Association for Machine Translation
in the Americas
PMB 300
1201 Pennsylvania Avenue, N.W., Suite 300
Washington, DC 2004 USA

Last name(s): _____ First name(s): _____ Title: _____

Address: _____

Home tel.: _____ Work tel.: _____ Fax: _____

E-mail: _____ Website: _____

Affiliation: _____

Professional associations: _____

Area of specialization:

MT User MT Developer MT Researcher Translator Manager Other _____

Method of Payment

- Check enclosed
 Credit card

Type of credit card: Visa MasterCard American Express

Card number _____ Exp. Date: ___/___

European Association for Machine Translation

APPLICATION FOR MEMBERSHIP

Please return this form, together with your payment or credit card information, to:

EAMT Secretariat, c/o TIM / ISSCO
 Université de Genève
 École de Traduction et d'Interprétation
 40, blvd du Pont-d'Arve
 CH-1211 Geneva 4, Switzerland

Type of member and membership fee per calendar year:

Individual SFr 50

Non-profit-making institution SFr 175

Representative: _____

Profit-making institution SFr 350

Representative: _____

Last name(s): _____ First name(s): _____ Title: _____

Address: _____

Home tel.: _____ Work tel.: _____ Fax: _____

E-mail: _____ Website: _____

Institution / organization: _____

Area of specialization:

MT User MT Developer MT Researcher Translator Manager Other _____

Method of Payment

Cheque payable to EAMT, enclosed

Banker's draft (copy enclosed) to account no. 351.091.40L

Union Bank of Switzerland
 Bahnhofstrasse 45
 CH-8021 Zürich, Switzerland

Please note: All bank charges must be borne by the applicant

Credit card (please provide information at right)

Type of credit card: Visa Eurocard

Card number _____

Expiration date: ___/___