MT News International

Newsletter of the International Association for Machine Translation

MT Summit VII Special Issue

Editors' note . . .

This issue of MTNI will be distributed at MT Summit VII in Singapore. It features items of special interest to the attendees, particularly the members of the IAMT associations.

MTNI is back on schedule. The following deadlines for submissions have been set for future issues:

#24 (vol. 8, no. 3): Dec 15, 1999
#25 (vol. 9, no. 1): May 1, 2000
#26 (vol. 9, no. 2): Sept 1, 2000
#27 (vol. 9, no. 3): Dec 15, 2000

As always, we welcome contributions and offers of assistance!

—MV and CB

Spotlight on the News

DARPA Announces Major Foreign Language Initiative: "TIDES"

Mindful of the burgeoning growth of non-English sources on the Internet, the U.S. Defense Applied Research Projects Agency (DARPA) has announced a five-year program to facilitate access to this material in English. The Translingual Information Detection, Extraction, and Summarization (TIDES) program is planned as a five-year DARPA undertaking that will develop the technology to seek, acquire, and utilize foreign language materials using English language tools. Foreign language materials on the Internet are now growing at such a rapid pace that they are expected to match or exceed the volume of English resources by the end of 1999. The TIDES program has been mobilized to help make these materials accessible and usable in English.

General Objectives

The overall intention is to develop, quickly and accurately, a comprehensive understanding of unfolding international situations in order to provide timely and valid interpretations to strategic and tactical decision-makers.

More specifically, TIDES intends to provide the capability of expressing an information need in English and using this formulated query to search among materials in a multitude of languages of interest to the U.S. Department of Defense (DoD). It will then:

- retrieve the relevant materials;
- translate their content into English;
- extract the names of people, places, organizations, and related entities;
- identify events of interest; and
- correlate the content of an array of documents occurring in multiple languages.

Natural language interfaces will be used. The information may be either

. . . TIDES from previous page

structured (tables, charts, diagrams, published text) or unstructured (informal written text, transcriptions of speeches, e-mail, etc.).

The goal of TIDES is to dramatically reduce the amount of time it takes to perform cross-language retrieval; information extraction, summarization, and interpretation; and machine translation of "new" languages.

Technical Topic Areas

For the first three-year period, research proposals have been solicited* and are currently under review in the following technical topic areas:

Machine translation. Application of corpus-based analysis or other ground-breaking techniques to the problem of rapid development of machine translation for new languages of interest. DARPA will select several dozen DoD-relevant languages for the demonstration of query translation, document translation, and query refinement using relevance feedback. "Query translation" is defined as the formulation of a foreign-language query based on English input; "query refinement," as the narrowing, broadening, or elaboration of the terms of a query in order to account for linguistic content.

Information access and management systems. Testing of the hypothesis that information access and management can be substantially enhanced by interactions between otherwise serial components. Examples of such interactions are topic detection and interaction, name and event correlation, and multi-document summarization. Interest extends to interfaces for multi-modal queries over high bandwidth connections with large, high-volume multimedia and multilingual data. "Topic detection" is the process of separating retrieved materials from multiple languages into categories of interest. "Name and event correlation" includes the detailed analysis of available resources in order to identify names of people, corporations and organizations, dates, times, events, and

related entities, and in order to establish correlations between related entities. The volume of information relevant to a specific problem can often overwhelm the analyst. Summarization typically strives to reduce the set of information to be examined by a factor of 10. Coherent summarization of multiple documents can also be used to sharpen the query and improve the return from information retrieval.

The TIDES program also intends to improve the performance of cross-lingual or translingual retrieval by exploiting the knowledge extracted in latter stages of the information access process, and by feeding this knowledge back to the user for refinement and back to the retrieval process as a more precise representation of the query. TIDES will test a number of hypotheses in this area. For example:

- Could machine translation be used to refine a foreign language query and improve retrieval performance by 50%?

- Could consistently occurring topics in the retrieved materials be identified and fed back through the retrieval process to improve performance by another 25%?

- Could names, places, events, and related entities be extracted from the results of a multilingual search,

- Could a coherent multi-document summary be used as a refined query to deliver translingual performance comparable to that achieved currently in monolingual systems?

Tools and resources. Development of corpus analysis tools (e.g., statistical, linguistic, inter-lingual based, hybrid) that permit the automated extraction of grammar and vocabulary from parallel and comparable corpora in multiple languages. Development of tools for rapidly building parallel and comparable corpora, as well as related linguistic tools for the rapid analysis of language. Development of annotated multilingual spoken and written language resources in support of the previous two technical topic areas. Development of architectures for modular approaches to provide interoperability of components.

---

*The deadline was July 26, 1999.

Goals of the Program

TIDES expects to meet the following targets by the end of the first three years:

• Performance on translingual queries at 75% of the performance for monolingual queries;

• Ability to establish a rudimentary query capability in a new language of interest within one month;

• Ability to establish a mature machine translation capability in a new language of interest within 12 months;

• Fifty percent accuracy in recognition of topics across multiple languages;

• Fifty percent accuracy in extracting named entities from documents in multiple languages;

• Generation of coherent, query-specific summaries from multiple documents in at least two languages.

In five years, TIDES intends to deliver the following:

• Translingual capabilities for at least 30 languages;

• Eighty percent accuracy in translin-

gual entity correlation;

• Seventy percent accuracy in filling out mutlilingual templates;

• Ability to generate coherent query-specific summaries of up to 20 documents in at least four languages.

Evaluation Criteria

Metrics-based evaluation will be used to assess progress. These tasks range from integrative applications to experimental one-time evaluations. Integrative applications include question-and-answer, summarization within and across documents, and database/template-filling.

It is also expected to evaluate technical components, including entity recognition and fact extraction. Experimental tasks include co-reference resolution and word sense disambiguation. Integrative applications are the main, overall tasks and will eventually be required as a benchmark for the program.

Further information: www.darpa. mil/ito/Solicitations.html. ❑

e-Lingo + Yupi.com Announce Services For Spanish Speakers

[Based on press release, 08/11/99]

e-lingo, a provider of Internet-based translation solutions, and Yupi. com, one of the largest Spanish-language portals worldwide, have announced that Yupi.com will begin offering e-lingo's translated text and browse services to its users.

Thomas Lamar, founder and CEO of e-lingo, explained that the Latin American on-line population is expected to grow by 1600% in the next two and a half years, but a large portion of the Web's content will still be being published in English. "Since Yupi.com is one of the largest portals serving this market," he said, "the alliance immediately makes e-lingo a leading provider of Internet-based translation solutions for this growing force of Web users."

The rapid expansion of the non-

SYSTRAN / Berlitz Launch Fykk Translation Services on the Internet

[Based on press release, 06/15/99]

SYSTRAN Software, Inc. has entered into a strategic alliance with Berlitz International, Inc., to offer professional human translation services on the Internet in addition to its existing machine translation services. The arrangement brings together SYSTRAN's MT technology and Berlitz's GlobalNET, a worldwide network of 22 production centers that provide translation and localization services in 19 countries.

SYSTRAN has been offering free on-line machine translation service from its Website since 1996. SYSTRAN's technology also powers the translation services on AltaVista and the Infoseek GO

Network. The blending of SYSTRAN's MT technology with expert human translation skills now offers significant added value for Internet users.

Speaking at the launch of the new service, Brian Kelly, vice president for Berlitz's European Operations, said: "We believe there is a lot of demand in the marketplace for translation activity performed quickly and at a lower cost. The new service will address both of these issues in a highly user-friendly environment."

Dimitrios Sabatakakis, chairman and CEO of SYSTRAN, went on to point out that "translation has become an everyday utility needed for Web-based and Intranet platforms." The alliance with Berlitz, he explained, allows SYSTRAN to provide its Internet and corporate clients with a professional service that is accessible 24 hours a day, seven days a week.

About Berlitz

GlobalNET is only a part of the larger Berlitz organization, which provides various language services through its 400 locations in 50 countries. In addition to direct services, Berlitz publishes a wide range of products including dictionaries, phrase books, travel guides, and self-study language instruction materials. In the last 12 months the company has acquired several small technology-based translation companies and has launched distance-learning and other technology-based learning initiatives.

Further information: SYSTRAN Software, Inc., tel.: (858) 459-6700; Website: www. systransoft.com, or Michael Gavin, Director, Sales and Marketing, Berlitz International, Inc.; tel.: +353 (1) 202 12 00; e-mail: michael.gavin@berlitz.ie; Website: www.berlitz.com. ❑

English-speaking Internet user population is creating a demand for a new layer of infrastructure—namely, real-time machine translation solutions for portals, Internet service providers, e-mail providers, e-commerce, and content sites. At the same time, the English-speaking world is seeking ways to access content on non-English Web sites (see "TIDES" on p. 1). According to TechServer, non-English Websites is expected to represent over half the content published on the Web by 2003.

To meet this challenge, e-lingo provides high-quality, real-time translation solutions, including Web-based search, text, e-mail, and browsing. On-line portals like Yupi.com, can incorporate e-lingo's translation solutions into their own Websites to offer their users greater service and greater access to on-line content.

San Mateo-based e-lingo is an Internet infrastructure company that develops and hosts private-label, real-time language translation solutions. These solutions include translated search, text, e-mail, and Web browsing.

Yupi.com was founded in 1996 in response to the need for high-quality Spanish-language content on the Web. All its content is categorized manually; it does not utilize web crawlers or spiders. Yupi.com and its network of sites is headquartered in Miami Beach and currently has offices in Argentina, Colombia, Mexico, and Spain.

Visit e-lingo at www.e-lingo.com; Yupi.com at www.yupi.com. ❏

ALPNET Acquires EP Electronic Publishing

ALPNET Inc., a translation company headquartered in Salt Lake City, UT, recently acquired EP Electronic Publishing Partners GmbH in Nuremberg, Germany. EP designs and develops custom-built systems for document, information, knowledge, and translation management.

In the past three years, EP has developed and implemented a translation management system. "EP's technology for translation, dictionary and content management is very advanced," said Jaap van der Meer, president/CEO of ALPNET.

Visit ALPNET at www.alpnet.com. ❏

Products

L&H Intros New Products for European Market

Belgian langtech superpower Lernout & Hauspie has introduced a new range of translation packages aimed at the European market. L&H Power Translator Pro is targeted at business users, while L&H Simply Translating is aimed at the home and educational markets. The initial languages supported (bidrectionally) are English, French, German, Spanish, Italian, and Portuguese.

Describing the differences between the two packages, L&H says Simply Translating is suitable for producing fast draft-quality letters, reports, e-mail messages, and more. Power Translator Pro, on the other hand, offers the business user additional features, including a range of thirty industry-specific dictionaries and support for Japanese as a source and target language. It also includes a Web page translator tool and facilities for customizing dictionaries. Both packages include a feature called Conversation Translation, which enables users to "chat" on-screen.

L&H Simply Translating will cost approximately EU$ 60, and L&H Power Translator Pro will run about EU$ 200.

Visit L&H at www.l&h.dom. ❏

Conferences

RIAO-2000: Content-Based Multimedia Information Access
Paris, April 12-14, 2000

Preliminary Announcement Call for Papers and Applications

Paris will be the site of the next in the RIAO triennial series of conferences on computer-assisted information retrieval, to be held on April 12-14. The encounter is being organized by the Centre de Hautes Etudes Internationales d'Informatique Documentaire (France) and the Center for the Advanced Study of Information Systems (USA), with the collaboration of a number of other organizations.

Past RIAO conferences have been held in Grenoble (1985), Boston (1988), Barcelona (1991), New York (1994), and Montreal (1977).

Under the general theme of "Content-Based Multimedia Information Access," topics will range from the traditional processing of text documents to the rapidly growing field of automatic indexing and retrieval of images and speech and, more generally, to all processing of audiovisual and multimedia information on various supports, including the Internet.

The conference is of interest for several scientific communities, including Information Retrieval, Natural Language Processing, Spoken Language Processing, Computer Vision, Human-Computer Interaction, and Digital Libraries. It is intended to serve as a forum for synergetic initiatives and forward-looking applications.

RIAO-2000 will present recent advances in scientific research, demonstrations of

Important Dates	
Proposal submission deadline:	November 1, 1999
Notification of acceptance:	December 15, 1999
Deadline for final papers:	January 15, 2000

prototypes resulting from this research, as well as the most innovative products now appearing on the market.

The Scientific Committee will select the papers, and the Applications Committee will accept proposals for presentation at the conference in the areas of document processing, information retrieval, spoken language processing, natural language processing, image processing, system architecture, and usage and best practice.

Within the area of NLP, the following topics are of specific interest: • Information extraction • NLP techniques for document processing • Terminology extraction and analysis • Automatic thesaurus construction • Multilingual and cross-lingual document handling • Machine translation of documents • Automatic summarization • Identification of the language of a text.

Further Information

Website: http://host.limsi.fr/RIAO. Contact: e-mail: riao2000@limsi.fr. Centre de Hautes Etudes Internationales d'Informatique Documentaires, 36-bis rue Ballu, 75009 Paris, France. Tel.: +33 (0) 1 42 85 04 75; fax: +33 (0) 1 48 78 49 61 or 1 45 26 84 45. ❏

Those interested in proposing a paper or application demo should fill out the Attendance Intention Form available on the conference Website, http://host.limsi.fr/RIAO.

---

to highlight similarities and downplay artificial distinctions.

A separate call for student papers will be announced soon. The same schedule for submissions will apply.

ANLP

The ANLP Program Committee, under the chairmanship of Sergei Nirenburg, New Mexico State University, invites papers describing natural language processing systems in the areas mentioned. In particular, submissions should be directed to mono- and multilingual text processing systems; spoken language and multimodal systems; integrated NLP systems; tools and resources for developing NLP systems; or the evaluation of performance of complete NLP systems.

NAACL

The NAACL Program Committee, chaired by Janyce Wiebe of New Mexico State University, invites papers on methodology, approaches, algorithms, models, analyses and experiments in computational linguistics. Program subcommittees are organized around eight main areas.

For the general sessions, papers are invited on substantial, original, and unpublished research contributions on all aspects of computational linguistics methodology, enabling technologies, approaches, algorithms, models, analyses, and experiments.

Further Information

Website available October 1. Questions regarding ANLP submissions should be addressed to: anlp2000-program@nmsu. edu, or: Program Chair, ANLP 2000, Computing Research Laboratory, Box 30001/Dept 3CRL, Las Cruces, NM 88003-8001. Questions with regard to NAACL submissions should be sent to: naacl2000-program@nmsu.edu, or: Program Chair, NAACL 2000, at the same address. ❏

Joint Conference: Applied Natural Language Processing and North American Chapter of ACL

Seattle, April 29 – May 2, 2000

Call for Papers

The Association for Computational Linguistics (ACL) is pleased to announce that the Sixth Applied Natural Language Processing Conference (ANLP) and the first conference of the new North American Chapter of the ACL (NAACL) will be held jointly on April 29 – May 3, 2000, in Seattle, Washington. The General Conference Chair is Marie Meteer, BBN Technologies.

The joint conferences will offer a unique opportunity to bring industry and researchers together to explore the full spectrum of computational linguistics and natural language processing, from theory and methodology to their application in commercial software.

For the general sessions, substantial, original, and unpublished contributions to computational linguistics are solicited. The ANLP program committee invites papers describing natural language processing systems—their development, integration, adaptation, and standardization; tools, techniques, and resources contributing to the development of complete end-to-end applications of NLP; evalua-

tion of system performance and related issues.

There is some inevitable overlap between the topic areas for NAACL and ANLP. In deciding whether to submit their papers to NAACL or ANLP, authors should consider whether their paper focuses more on the methodology or the end application of that methodology to solve a particular problem. We also appreciate that it can be advantageous to view the same work from both a theoretical/methodological perspective and an applied perspective; we welcome paired submissions to NAACL and ANLP, though each submission needs to make a significant contribution on its own. Papers will not be exchanged between the two program committees. However, in the final program, papers may be grouped or juxtaposed in related sessions

Important Dates	
Submission notification deadline	November 10, 1999
Paper submission deadline	November 17, 1999
Notification of acceptance of papers	February 1, 2000
Deadline for camera-ready papers	March 12, 2000

CLAW-2000:
3rd International Workshop On Controlled Language Applications

Seattle, April 29-30, 2000

The 3rd International Workshop on Controlled Language Applications will be held on April 29-30, 2000, as a two-day pre-conference workshop in conjunction with the joint meeting of Applied Natural Language Processing (ANLP) and the North American Computational Linguistics Association (NACLA) in Seattle, Washington, at the Westin Hotel.

The two-day workshop will feature invited speakers, paper presentations, poster sessions, panel discussions, and system demonstrations.

Topics of Interest

Of interest are submissions on all topics relating to the design, implementation and application of controlled languages, including but not limited to: human communication protocols; controlled text authoring; conformance checking systems; controlled language authoring aids; machine translation; memory-based translation authoring and translation; issues in Controlled Language design; industrial experience and evolving requirements; quality and usability evaluations of controlled language.

On the research front, we are grappling with issues like standardization, design, and evaluation of controlled language systems. On implementation issues, we are particularly interested in studies regarding author productivity and document usability. The workshop will give equal emphasis to the academic, corporate and industrial perspectives, while bringing together researchers, developers, users, and potential users of controlled language systems from around the world.

Further Information

For further details on submissions, see Website, www.up.univ-mrs.fr/~veronis/claw2000. For general inquiries, contact Jeff Allen, e-mail: postediting@hotmail.com. ❑

Important Dates	
Proposal submission deadline:	November 30, 1999
Notification of acceptance:	January 20, 1999
Deadline for camera-ready copy:	March 1, 2000

LREC-2000: 2nd International Conference On Language Resources and Evaluation

Athens, May 31-June 2, 2000

Announcement and Call for Papers

The European Language Resources Association (ELRA), the Institute for Language and Speech Processing (ILSP, Athens, Greece), and the National Technical University of Athens, Greece, are pleased to convene the Second International Conference on Language Resources and Evaluation, to be held in Athens on May 31-June 2, 2000. The Conference Chair is Antonio Zampolli, of the University of Pisa, Italy.

The Scientific Programme will include invited talks, presentations of accepted papers, poster sessions, referenced demonstrations, and panels. Pre-conference workshops will be organised on the May 29-30, and post-conference workshops on the June 3-4, 2000.

Conference Aims

The aim of this conference is to provide an overview of the state of the art, discuss problems and opportunities, and exchange information regarding ongoing and planned activities that involve language resources and their applications.

In the framework of the Information Society, the pervasive character of human language technologies has been widely recognised. Two issues are currently considered to be particularly relevant:

* availability of language resources; and
* methods for the evaluation of resources, technologies, and products.

The term *language resources* refers to sets of language data and descriptions in machine-readable form used specifically for the building and evaluation of natural language and speech algorithms or systems; for software localisation industries and language services; for language-enabled information and communication services; and for electronic commerce, electronic publishing, language studies, subject-area specialists, and end users. Examples of language resources are written and spoken corpora, computational lexicons, grammars, terminology databases, and basic software tools for the acquisition, preparation, collection, management, customisation, and use of these and other resources.

The relevance of evaluation for language engineering is increasingly recognised. This involves assessment of the state of the art for a given technology, measuring the progress achieved within a programme, comparing different approaches to a given problem and choosing the best solution, knowing its advantages and drawbacks, assessing the availability of technologies for a given application, benchmarking products, and assessing user satisfaction.

Language engineering and R&D in language technologies have made important advances in the recent past in various aspects of both written and spoken language processing. Although the evaluation paradigm has been studied and used in large national and international programmes—including the US (D)ARPA program, the EU Language Engineering programme, the Francophone Aupelf-Uref, and others—there are still unresolved problems that call for basic research.

The conference topics will fall under the following three areas:

- Issues in the design, construction, and use of language resources (theoretical and best practice);
- Issues in the evaluation of human language technologies; and
- General issues.

Further information

General information: LREC Secretariat: Despina Scutari, ILSP, 6 Artemidos & Epidavrou Str., 15125 Marousi, Athens, Greece. Tel: +30 (1) 6800959; fax: +30 (1) 6854270; e-mail: LREC2000 @ilsp.gr; LREC-2000 Website: www.icp.grenet.fr/ELRA/ lrec2000.html. ❑

Important Dates

Proposal submission deadline:	November 20, 1999
Notification of acceptance of workshop and panel proposals:	December 15, 1999
Notification of acceptance of papers, posters, referenced demos:	February 2, 2000
Deadline for articles to appear in proceedings:	April 2, 2000

Users' Forum

SDLX

Patrick Powell
i2 Technologies

SDLX is a globalization productivity tool which allows the re-use of common translated terms and phrases, thereby increasing translator productivity.

As a beta tester for SDLX, I was able to constructively use much of the offered functionality that this suite of tools encompasses for various live projects.

i2 Technologies produces various client/server applications in the supply chain, ERP, and logistics areas. Due to their size, these applications normally have extensive hardware and software requirements. A typical implementation of a given i2 product at a customer site can easily take up to a year or more. Localizing these applications can prove to be a daunting task, especially because terminology must remain consistent over the entire given range of products and on all supported platforms (Windows, Unix). In addition, many of these programs do not follow the standard coding practice of creating language-specific .dll or .rc files, relying instead on proprietary techniques to relay the translated client and server strings to the user interface.

SDLX proved to be a convenient and sta-

ble tool in handling our various file format requirements. The three prevalent file types which I used were .rtf, .txt, and .rc files. The languages included French, German, Hungarian, and Japanese. My two initial goals were (1) to apply existing old translations that were still in their native file formats to new updated versions, and (2) to create master glossaries which would take into account terminology from all products across all platforms.

Installation of SDLX went without a hitch. Its performance on a Japanese Windows 95 machine was entirely as expected. Japanese characters in translated texts were displayed correctly throughout. Since I already had existing translations for several products and languages, I needed to align the original English terminology with the respective translations.

My experience with other tools such as Trados had prepared me to expect tedious work ahead, but SDLX's stand-alone application format proved to be surprisingly user-friendly. Alignment of .rtf, .txt, and .rc files (which were opened in standard text mode) went more smoothly than expected. Occasional misalignments could be manually fixed within seconds. The result could then be exported in various database-friendly formats.

With the help of these glossary data-

bases, I was able to create appropriate master glossaries in the .csv format, one per language. The .csv format was convenient for further processing using specific Perl scripts that can query the file via a Web browser and return the appropriate information in real time. We used this to create an on-line translation database for our intranet, an enhancement which has already proven extremely helpful in resolving quick translation questions—for instance, a question from a consultant in Germany at 4 a.m. in the United States.

The SDL Edit module was very easy to use. I once again preferred the stand-alone application format, as compared with Trados' integration with WinWord.

The tool is set up in such a way that a side-by-side comparison of the original text and the suggested translations is possible. Standard editing and searching features are also supported. In addition, a match value (expressed as a percentage) is given for the source vs. suggested translations, along with the option of displaying further suggestions for the translation.

The source file can be "seeded" with the suggested translations according to a user-selectable match value threshold, thus leaving only truly new text to be translated. I used this feature for several .txt and .rtf files with great success. Since SDLX can convert the listed files (as well as .mif and html) into the Open Tag format, I was able to use translations across all file types without being restricted to one particular file type per translation.

Given SDLX's inherent stability and convenience, I would recommend it to anyone requiring a solid translation memory tool.

SDLX includes more features that I have not elaborated on here. Examples are its ability to maintain special formatting within the files and its ability to import other translation memory databases—e.g., the TMX, delimited, and Trados formats. i2 will continue to use this tool regularly.

SDLX 2.0 will be sold on-line at http://sdlx. sdlintl.com. ❑

Association News

Agenda, Fifth General Assembly
International Association for Machine Translation (IAMT)

National University of Singapore, Kent Ridge Digital Laboratories
Singapore, Thursday, September 16, 1999, 1:00 – 2:00 p.m.

1. Call to order (*H. Tanaka*)

2. Adoption of the agenda (*H. Tanaka*)

3. Approval of the minutes of the 4th General Assembly (*H. Tanaka*)

4. Report of the President (*H. Tanaka*)

5. Report of the Secretary (*S. Talbott for J. White*)

6. Report of the Treasurer, budget for 1999 (*S. Talbott for R. Merchant*)

7. Report of the Editor, *MT News International* (*M. Vasconcellos*)

8. Report of AAMT (*H. Tanaka*)

9. Report of AMTA (*E. Hovy*)

10. Report of EAMT (*V. Hansen for J. Hutchins*)

11. Actions of the 7th Council (*S. Talbott*)

12. New business (*H. Tanaka*)

13. Presentation and confirmation of officers (*H. Tanaka*)

14. Statement of the new President (*B. Maegaard for J. Hutchins*)

15. Date and place of the next General Assembly (*B. Maegaard for J. Hutchins*)

16. Adjournment ❏

Address by W. John Hutchins

On Assuming the Presidency of IAMT

MT Summit VII, Singapore, September 16, 1999*

I would like to take this opportunity to review what the IAMT has achieved, and what it could be doing in the next few years. The purpose of the International Association for Machine Translation, as set forth in its Articles of Incorporation, is to "bring together users, developers, researchers, sponsors, and other individuals or institutional or corporate entities interested in machine translation for the purpose of studying, evaluating, and understanding the science of machine translation and educating the public on important scientific techniques and principles of machine translation." For this purpose, five main areas of activity are envisaged.

Dissemination of Information

The first of the areas of activity is the "sharing of knowledge" about MT "through the collection, compilation, exchange, and dissemination of information." The principal organ in this area of activity has been the IAMT newsletter, *MT News International*. Although intended primarily for members of IAMT and its regional associations, it is also seen, and subscribed to, by non-members and has served to attract new members.

*Read by Bente Maegaard

MTNI was established both as a channel for keeping members informed about the work of their associations, in particular past and future conferences and workshops, and as a medium for reflecting and reporting on world-wide activity in general within the field of MT and related areas. It has largely succeeded in this aim, but what has remained a disappointment for me—and here I speak as a previous editor—is that there have been too few reports from the users of MT systems. The reasons are easy enough to understand: users are busy people; MT is just one aspect of their work; and perhaps they are reluctant to discuss whether their investments in MT systems have been successful or not. Whatever the reasons, *MTNI* would like to hear more from actual users.

This year another publication has been launched with the aim of disseminating information. This is the *Compendium of Translation Software*, published by the European Association for Machine Translation and distributed in draft form to all IAMT members. The intention is to produce a first edition for sale to the general public sometime early next year. The aim of the Compendium is to provide, in a reasonably concise and accessible form, information about currently available products—not just MT systems (for business and home use), but also the wide range of computerised translation tools (translation workstations, terminology management systems, etc.). It aims to list what is now available and at what prices, what languages are covered, which systems are suitable for large-scale use, which for use on the Internet, and so forth.

It is particularly important that the Compendium is as accurate and up to date as possible. If you have not done so already, please send information about systems and software for inclusion in the forthcoming printed edition.

One question to be resolved in the near

future is whether the Compendium should continue to be a printed publication or whether it should be only in electronic form. The EAMT is currently mounting the Compendium on its Website, and it is also investigating the possible creation of an electronic database similar in content to the Compendium which would be easier to update, more accessible, and easier to search. The opinions of the membership will be welcomed.

Organisation of Conferences and Workshops

The second area of IAMT activity is "sponsoring and supporting workshops, symposia, and conferences on machine translation and related technologies and applications." In this area, IAMT can point to significant successes. IAMT's own series of conferences, the biennial MT Summits, are well established as the major forum for all those involved and interested in MT—users, developers, vendors, and researchers.

> *There have been too few reports from the users of MT systems*

The regional associations have been equally successful. AAMT has held regular conferences and workshops, both large and small, in its region. EAMT has held annual workshops which bring together vendors, developers, users, and potential users in fruitful exchanges of views and experience. AMTA has organised three major conferences alternating with the MT Summits, which have attracted large numbers from North America and elsewhere. These conferences successfully combine reports of research and development, accounts of users' experiences, demonstrations of systems, and large exhibitions of MT and related products. All three associations maintain close relationships in their regions with other national and international organisations in related fields.

Establishment of Criteria

A third area of IAMT activity is "discussing and establishing reference criteria for the evaluation of machine translation technology." In this area, it is only recently that IAMT has begun to be active. Last year a group within IAMT began an attempt to draw up categories and specifications of different types of MT systems and translation aids, with the aim of establishing some mechanism for the labelling or 'certification' of products. It must be stressed that certification does not imply any evaluation of software; it would simply be a rough guide for potential purchasers to help them to identify whether particular software might or might not do what they expect.

Large organisations can usually afford to make their own assessments of software, but small businesses and individual members of the general public can be bewildered by the range of products on offer and can be easily misled into buying something that is not appropriate for their needs.

At this stage, it is not clear whether IAMT certification would receive the approval of vendors or members, nor is it clear how the process itself would be undertaken—whether it would be done by a group within IAMT, by an independent outside body, or by a group of vendors and users. Shortly, proposals will be put forward for comment.

> *Large numbers of the general public are becoming aware of the existence of MT systems and translation software*

It is also not clear whether certification should be a first tentative step towards "reference criteria for the evaluation of machine translation

technology." This does not mean setting up a body to undertake the evaluation of systems. What it suggests is that IAMT could establish criteria that organisations could apply themselves when they do their own evaluations. In this area, valuable efforts have already been made in Japan by JEIDA and AAMT. As more and more organisations are investigating the use of translation software, I believe the time has come for IAMT itself to draw up guidelines for the evaluation of systems. For this, it ought to be able to draw upon the experience of many of its own members.

In this connection, however, I believe we could, as a community, go further. One possibility is that the IAMT could encourage the formation of networks of MT consultants. There are many people within companies and associations with experience in the practical day-to-day use of MT systems and translations tools. Some of them would probably be willing to offer their expertise and advice to others new to the field—obviously at mutually agreed fees. The IAMT or one of the regional associations could act as a clearinghouse, putting enquirers into touch with consultants and advisers.

Access to Multilingual Corpora

The fourth area of activity has not so far been undertaken by IAMT at all. That is: "facilitating access by researchers to machine-readable corpora and co-operating in the exchange of formats and text-encoding conventions." The principal reason is that other organisations have been already active in this area. I need only mention the European Language Resources Association (ELRA), which has been co-ordinating access to lexical and text corpora with great suc-

cess; the Localisation Industry Standards Association (LISA), which is particularly involved in the standardisation of exchange formats and text-encoding protocols; and the Linguistic Data Consortium (LDC), which for a number of years has been archiving and making available many valuable multilingual text and speech corpora. IAMT and its regional associations should keep a close eye on developments in order that the interests of MT developers and users are properly represented. Closer involvement may be desirable. What are the views of members?

A loosely related issue is what links IAMT should have with associations and organisations involved in the development of multilingual software. Increasingly, systems for database searching, information extraction, summarisation, etc., embrace cross-language and translation components. This conference will have heard reports of this activity. It is to be expected that in the future MT systems will be available not just as standalone programs but also as integral parts of other natural language tools. IAMT cannot afford to ignore these developments. The views of members on this issue will be welcomed.

Educational Materials and Programs

The fifth and last area of activity set down for IAMT is the somewhat vague and all-embracing one of "developing appropriate educational materials and programs." Here, too, IAMT has not yet made a mark. However, it is perhaps an area which IAMT together with its regional associations should address urgently. As we all know, large numbers of the general public are becoming aware of the existence of MT systems

and translation software. Public exposure through Internet services, software for Web translations, cheap PC-based MT programs, etc., is greater than anyone might have expected even a few years ago. Indeed, for many people using Internet resources, MT is more than an interesting sideline; it is fast becoming a necessity.

At the same time, there is much ignorance about what MT is and what it is not, what it can do and what it cannot do. There is a widespread assumption that MT systems are intended to produce 'real' translations—the kinds of translations that human translators produce. This is not surprising, since the marketing of MT systems often holds out such expectations. When the systems do not produce 'real' translations, the general public feels cheated.

What we need to get over to the general public is that translation systems provide rough versions which give users assistance in their understanding of foreign language texts, which can help them to communicate in foreign languages, which can help them to communicate in foreign languages, and which can give basic drafts for human revisers to produce higher-quality translations. MT systems should be judged according to their usefulness and cost-effectiveness, as aids for cross-language communication, and not as substitutes for human translators. The certification initiative and the Compendium may help in this education process, but only to a limited extent.

We have to change the perceptions and prejudices of those who have an influence in moulding public opinion: the authors of popular science articles, technical writers, journalists, broadcasters, and so forth.

> *We have to change the perceptions and prejudices of those who have an influence in moulding public opinion*

- Perhaps IAMT and its regional associations could produce a leaflet or pamphlet or Web page presenting MT and translation aids in straightforward non-technical language.
- Perhaps representatives of IAMT and its associations should be designated to attend and speak at meetings and conferences where translation and cross-language communication are likely to be topics of discussion.
- Perhaps IAMT should sponsor events specifically aimed for the general public and for the various media organisations.

And there must be other possibilities.

In Conclusion

I have put forward some suggestions for how the IAMT might go further in fulfilling the aims of its founders. Briefly, these include the Compendium and an electronic database, the certification of productions, guidelines for evaluation, networks of consultants, wider links in the area of multilingual corpora and software, and the development of various educational materials and programs.

Some of these aims are achievable in the near future, and I hope they will be. Others are more long-term objectives. However, we do not have to be restricted to the five areas of activity mentioned in the IAMT Articles of Incorporation. As just one example, it has been proposed that IAMT should set up an archive for the collection of historical materials related to machine translation over the past fifty years—and this is something that I have a personal interest in.

I am sure that members here and elsewhere will have many other ideas for what IAMT and its regional associations could and should be doing—within the inevitable limitations of a wholly voluntary non-profit organisation. IAMT has succeeded in its first objectives to bring together users, developers and researchers. It is time for us to look to new challenges. ❏

EAMT General Assembly

Krystal Hotel, Prague
April 22, 1999

The Assembly was called to order at 4:45 p.m. by EAMT President, *John Hutchins*, who chaired the meeting.

The following members were present: Colin Brace, Anthony Clarke, Daniel Grasmick, Viggo Hansen, John Hutchins, Jaro Lajovic, Elena Paskaleva, Bente Maegaard, Carlos Paz, Dimitri Theologitis, Anne Tucker, Grazyna Vetulani, Zygmunt Vetulani.

1. Approval of the Agenda

The Chairman reviewed the agenda, which the Assembly then approved unanimously.

2. Executive Committee Reports

President's report. The President received the approval of the Assembly that agenda item 4 (Report on Activities) be taken as part of the President's report.

In his report, the President mentioned that the Geneva EAMT Workshop had been held in April 1998. The President further reported on the *Compendium of Translation Software*, a preliminary draft edition of which was now available. The President encouraged EAMT members to report to him any amendments or corrections for inclusion in the forthcoming Compendium.

Secretary's report. The Secretary (*Viggo Hansen*) informed the Assembly of the membership status. By the end of 1998 EAMT had 82 individual members, 11 non-profit making institutions, and 13 commercial companies. The comparable figures for the end of 1997 were 78, 6, and 9, respectively.

Treasurer's report. The Treasurer (*Anthony Clarke*) presented the 1998 EAMT accounts. (Exhibit 'A'). The organisation is in a financial healthy condition, having assets totalling approximately SwFr 41,000.* The Treasurer reminded the Assembly of the fact that EAMT will be responsible for MT Summit VIII in the year 2001, which will require a sound economy.

The Treasurer encouraged members to recruit new members to the organisation, in particular new corporate members. The Treasurer's report was approved by the General Assembly.

3. Newsletter Report

MTNI Co-editor *Colin Brace* gave a report on the newsletter situation. There have been problems to get out the newsletter to the European members. He reported that things had recently begun to improve. Since October 1997 Muriel Vasconcellos has taken over responsibility for editing the newsletter.

4. Report on Activities

This agenda item was dealt with under agenda item 2. President's report.

5. Report on Web Pages

Colin Brace gave information about the Web pages of EAMT, which can be found at *www.eamt. org.* There are about 300-400 visitors a month. There was discussion about providing access for members only to the *Compendium of Translation Software* and the proceedings of EAMT workshops. These suggestions will be considered by the EAMT Webmaster (Colin Brace). It was also proposed to have links to MT-related Websites. Members are encouraged to suggest any such link to Colin Brace (*eamt@cst.ku.dk*).

6. 2000 Membership Fees

The Treasurer proposed and the General Assembly agreed that the membership fees for 2000 should be remained unchanged, as follows:

Individual member, student	SwFr 25
Individual member	SwFr 50
Non-profit institution	SwFr 175
Company	SwFr 350

7. Election of Officers

Officers are elected for a three-year term. All members of the Executive Committee were elected last year and have served their first year of the three-year term. The members of the Executive Committee are: John Hutchins, President; Viggo Hansen, Secretary; Anthony Clarke, Treasurer; Colin Brace (MTNI Co-editor); Bente Maegaard, Dimitrios Theologitis, Jörg Schütz (*ex officio*).

8. Future Activities

The President mentioned the following future activities:

• TMI in Chester, England, August 22 – 25, 1999: EAMT will be responsible for a tutorial on August 22.

• MT Summit VII in Singapore, September 13-17, 1999.

• EAMT-2000: Venue and time not yet decided, but there is an offer from Ljubljana to host the workshop.

• MT-2000: International Machine Translation Conference in Exeter, England, November 20-22, 2000, organised by the British Computer Society.

• MT Summit VIII in Santiago de Compostela, Spain, September 18-22, 2001.

9. Other Business

The President mentioned the IAMT Award of honour. Any member of EAMT is entitled to nominate an individual to receive the award. Entry forms will be mailed directly to all EAMT members.

There being no other business, the President adjourned the meeting of the 1999 General Assembly.

Submitted by Viggo Hansen
EAMT Secretary ❏

---

*Full report available on request.

Compendium of Translation Software: Members Asked To Contribute

By now, all IAMT members should have received their free copies of the Compendium of commercial machine translation systems and computer-aided translation support tools. This is a *preliminary* edition; in December this year or January next we will published a true first edition, to be put on sale to the general public (and at a very low nominal price to members of the IAMT associations).

The aim of the Compendium is to provide a detailed, accurate, and up-to-date listing of all systems that are currently available for purchase on the market. It will not include any commercial systems still under development, or any systems or translation tools of limited availability, such as those developed for particular clients by software or other IT companies.

The Compendium should be seen as a publication by the *IAMT membership as a whole*, and it is particularly important that every member makes some contribution to its content, to ensure its accuracy and currency.

Members of the IAMT associations are strongly urged to look in the Compendium for any products and companies that they know about, to check whether they are listed, and, if they are, to check for any mistakes or omissions. *Please inform us of any errors or gaps (however small). Even if the information is correct, please let us know.* Full details about the coverage of the Compendium can be found in the preface and in the section entitled 'Arrangement of Entries'.

Information should be sent as soon as possible to the compiler, John Hutchins, e-mail address: WJHutchins @compuserve.com. ❑

Calendar

1999

September 13–17: MT Summit VII, Kent Ridge Digital Labs, National University of Singapore, Singapore. General Chair: Hozumi Tanaka; Local Organizing Committee Chair: Hwee Boon Low; Secretary: Victorine Chen-Toh; Program Committee Chair: Jun-ichi Tsujii; Vice Chair: Loong Cheong Tong. *Information: www.krdl.org.sg. Contact: Victorine Chen-Toh, e-mail: vicky@krdl.org.sg; Tel: +65 874-2003; fax: +65 776-8109.*

November 10-11: Translating and the Computer 21, London, UK. Organized by Aslib, the Association for Information Management, and supported by the Institute of Translators, the British Computer Society, the International Association for Machine Translation, and the European Association for Machine Translation. Program chairs: Daniel Grasmick, Ruslan Mitkov, and Chris Pyne. *Information: www.aslib.co.uk/conferences/tc20form. html;tel.: +44 ((0)20) 7903 0000; fax +44 ((0)20) 7903 0011. Contact: nicole. adamides@aslib.co.uk.*

2000

March 2-4: First International Conference on Specialized Translation, Barcelona, Spain. Organized by the Facultat de Tradicció i Interpretació and the Departament de Traducció i Filologia of the Universitat Pompeu Fabra. *Information: www.upf.es/dtf/ index.htm. Tel: +34 935 422 409 / +34 935 422 275; fax: +34 935 421 617; e-mail: congres.trades@grup. upf.es.*

April 12-14: *RIAO-2000: Content-Based Multimedia Information Acess, College de France, Paris. France. Organized by Centre de Hautes Etudes Internationales d'Informatique Documentaire and Center for the Advanced Study of Information

Systems. *Information: http://host.limsi. fr/RIAO. Contact: riao2000 @limsi.fr. Tel.: +33 (0) 1 42 85 04 75; fax: +33 (0) 1 48 78 49 61 or 1 45 26 84 45.*

April 29-30: CLAW-2000: 3rd International Workshop on Controlled Language Applications, Westin Hotel, Seattle, Washington, USA. *Information: www.up.univ-mrs.fr/~veronis/ claw2000. Contact: Jeff Allen, e-mail: postediting@hotmail.com. [*]*

April 29-May 2: Joint Conference of Applied Natural Language Processing and North American Chapter of the Association for Computational Linguistics, Westin Hotel, Seattle, Washington, USA. *Information: Website to be available October 1. Questions regarding ANLP submissions should be sent to: anlp2000-program@nmsu.edu. Inquiries with regard to NAACL 2000 should be addressed to: naacl2000-program@nmsu.edu. [*]*

May 31-June 2: LREC-2000: 2nd International Conference on Language Resources and Evaluation, Athens. Organized by the European Language Resources Association and the Institute for Language and Speech Processing. *Information: www.icp.grenet.fr/ELRA/ lrec2000.html. Contact: LREC2000 @ilsp.gr. Despina Scutari, LRED Secretariat, c/o ILSP. Tel: +30 (1) 680 0959; fax:+30 (1) 685 4270. [*]*

November 20-22: MT-2000: International Machine Translation Conference, Exeter, UK. Organized by Exeter University and Natural Language Translation Specialist Group of the British Computer Society. *Information: David Wigg, Chairman, BCS-NLTSG, wiggjd@sbu.ac.uk.*

2001

September 18-22: MT Summit VIII, Santiago de Compostela, Spain. ❑

---

*See Conferences section of this issue for further details.