MT News International

Newsletter of the International Association for Machine Translation

Published in February, June and October

ISSN 0965-5476 Issue no.16, February/March 1997

IN THIS ISSUE:

Editor-in-Chief:
John Hutchins, The Library, University of East Anglia, Norwich NR4 7TJ, United Kingdom
Fax: +44 1603 259490; Email: J.Hutchins@uea.ac.uk; or: 100113.1257@compuserve.com
Regional editors:
AMTA: David Clements, Globalink Inc., 4375 Jutland Drive, Suite 110, San Diego, CA 92117, USA. Tel: +1 619 490-3680 x 302; Email: 71530.3476@compuserve.com
EAMT: Jörg Schütz, Institute for Applied Information Sciences (IAI), Martin-Luther-Str.14, D-66111 Saarbrücken, Germany. Fax: +49 681 3895140; Email: joerg@iai.uni-sb.de
AAMT: Professor Hirosato Nomura, Kyushu Institute of Technology, Iizuka, 820 Japan.
Fax: +81 948 29-7601; Email: nomura@dumbo.ai.kyutech.ac.jp
Advertising Coordinator:
Bill Fry, Association for Machine Translation in the Americas, 2101 Crystal Plaza Arcade, Suite 390, Arlington, VA 22202-4616, USA.
Tel: +1 703 998-5708; Fax: +1 (703) 998-5709.

Published in the United States by Jane Zorrilla

SPOTLIGHT ON THE NEWS

Major Management Shake-Up at Globalink

The Board of Directors of Globalink, Inc. announced a series of management, organizational and cost-cutting measures implemented as of the close of business on January 6, 1997. Since Globalink is the leading manufacturer of retail MT software, this series of moves could have unknown but far-reaching impact on the MT industry.

Chairman & CEO Harry Hagerty will serve as acting President. Mr. Hagerty replaces Jim Lewis, who will become a professional consultant to the company. Lewis took over for Michael Tacelosky as company president in 1995.

These actions also involve the realignment of the duties of several key corporate officers. Philippe Kuperman, who was appointed Vice President of International Sales in August 1996, will

now assume increased responsibility worldwide as Executive Vice President of Sales & Marketing. Mr. Kuperman stated, "We anticipate substantial growth in revenue through new distribution opportunities worldwide as well as significant new OEM partnerships."

Brian Garr, who previously managed computer information, Internet and technical support activities, has been named Chief Technology Officer. As such, he manages the Research & Development departments of the company. This appointment coincides with the departure of Tim Meekhof from Globalink. Meekhof—chief architect of the much-publicized Barcelona system—has served as Chief Scientist and advisor to the President of Globalink since 1995. Meekhof plans to join with James Hicke, former Vice President of Research and Development at Globalink, in Chantemar Creations, a fledgling game software firm. Globalink has made no announcement about finding a replacement for the Chief Scientist position.

The linguistics department at Globalink is also losing one of its key members: Christine Fisher, Assistant Director of Linguistics. After leaving Globalink at the end of January, Fisher plans to pursue a career in training and technical writing. Ms. Fisher was recently at AMTA-96 in Montreal, where she demonstrated the Barcelona system for an audience of developers, translators and researchers.

The implementation of the company's restructuring activities is expected to result in an annualized savings of $2.8 million, and was -— according to the company -- initiated in areas other than those directly involved in revenue generation. Globalink's newly-appointed CFO, Mark Paiewonsky, commented, "These measures are not expected to have any negative impact on revenue; instead, they will result in greater operating profit contributions in future quarters."

In the area of technology development, Mr. Garr commented, "We are looking forward to capitalizing on our investment in our Barcelona technology, additional language development, client/server solutions and language learning titles." At this point, it remains unclear what changes the new management will make in the overall MT research strategy of the self-titled "Translation Company."

Systran PRO for Windows is Finalist for Awards

[Press release]
SYSTRAN Software Inc. (SSI) has been named one of the three finalists in the Software Division of the "Most Innovative New Products Awards" sponsored by UCSD Connect, a high technology enterpreneurship program associated with the University of California San Diego.

SSI is being honoured for its SYSTRAN PROfessional for Windows, a software program that translates ... 13 language combinations. Introduced in 1995, the program allows individuals and corporate users to access patented technology that was previously available only to the US intelligence community and the largest multinationals via mainframe computers.

SYSTRAN PROfessional for Windows preserves the format of the original document, which can be prepared using any of the leading word processor programs. SSI maintains large, multi-target language dictionaries that contain a total of more than 2.5 million entries in 16 different languages. The program has a user-friendly environment that allows customers to easily add their own terms and expressions. Language pairs available now include English from/to French, Italian, German, Spanish, Portuguese and Japanese; and Russian into English.

"Considering there are 350 software companies making an international impact in the San Diego region, we are particularly pleased with this recognition," said company president Denis Gachot. "Being named one of the three finalists is in itself quite an honor."

SSI also offers translation services for all languages and software development of new language pairs at its world headquarters in La Jolla, Calif. SSI's homepage is <http://www.systranmt.com>.

TranSmart wins Finnish Software Prize

At the beginning of November, TranSmart was awarded the Oiva prize. Oiva is an annual Finnish software contest in which the year's best software products are chosen. Innovative solutions, general applicability, technical quality, commercial prospects and user friendliness are some of the properties evaluated. This year the jury chose TranSmart, out of over 50 contestants, as the best Finnish software product of 1996. TranSmart was especially praised for its translation speed and quality.

CONFERENCE REPORTS

AMTA-96

Biennial AMTA Conference at Montreal

2-5 October 1996:

Two personal views

Laurie Gerber (Systran)

AMTA '96 was a remarkable moment in MT history. It included dramatic events, such as the transition to a new president. Muriel Vasconcellos, AMTA founder, stepped down after 6 years as the instigator and shaper of the organization, handing over control to newly-elected Eduard Hovy. It also included subtle evidence that AMTA has become a rather stable organization with a life and momentum of its own. Muriel seemed to surprise herself when she became emotional describing this conference as the high point of her career. The establishment of the AMTA conference as a regular event, and the strength of the association's leadership, mark the fruition of seeds she has planted. She expressed her satisfaction that she is able to step down as president with confidence. It appears that the organization she founded will continue in the same spirit of inclusiveness, exploration, and growth that has marked it from the beginning.

The conference was organized into a series of tutorials, talks, panel discussions, technical papers, and system demonstrations. In addition, a special room was set aside for exhibitions. The conference was held in the Hotel Radisson Gouverneurs in downtown Montreal, Canada. The location is very well set up for a conference the size of AMTA (ca.150 participants), with several conference rooms of varying sizes allowing 2 sessions to proceed at once, as well as offering a very pleasant "indoor terrace" for socialization. The unfortunate coincidence that the hotel was on strike did little to dampen enthusiasm and, thanks to Elliott Macklovitch's tireless efforts, did not hinder the proceedings.

The conference was preceded by two days of workshops. Tuesday, October 1 was the first event, a "pre-workshop-workshop" on interlinguas. Wednesday, October 2 included 6 well-organized half-day tutorials addressing the interests of researchers, developers, users, and potential users. Tutorial titles included: MT Evaluation; MT for Personal Use; Knowledge-Based, Example-Based and Multi-Engine MT; Statistics in MT; Choosing an MT System; and The Lexicon for MT. The workshops were very well attended. The half-day format allowed presenters to give solid, in-depth introductions to these topics. The response from attendees of the various sessions was very positive, indicating that the focused, in-depth sessions serve a real need.

In order to include more sessions with multiple speakers -generating a dialog between the speakers and audience - several sessions were conducted which included two speakers on the same or similar topics. The experiment was successful and provided a chance for each speaker to present their topic in some depth (45 minutes), while also giving the audience additional perspective via the back-to-back presentations of differing views or experiences from experts in the field. The program committee, chaired by Eduard Hovy of USC/ISI, developed a program that was well-varied and very interesting to attendees. The only complaint heard was that all the sessions were of interest, but one could only attend half of them. The answer to this complaint comes in the form of something new at this conference - videotaping of all of the sessions. These videotapes are available for purchase.

In discussing the conference with other attendees - especially those who needed to report back to their bosses and organizations - there was a summary comment, "I didn't really see anything new."

This is hardly surprising, as we are still digesting the introduction of the various corpus and knowledge based techniques which sparked so much interest in the early 1990s. Coming up with theories and proof-of-concept experimental systems is relatively quick and tidy. It's scaling up - building a large enough lexicon, gathering data, testing, debugging and refining - all of which are long-term, slow-moving endeavors - that takes up most of the time. In place of new or surprising developments, many presenters discussed the ways in which MT development is being hybridized. Research as well as commercial development is moving beyond the "proof-of-concept" stage to show repeatability, and general applicability of methodologies. Experiments are underway to determine the best combinations - which tools solve which problems best.

In the arena of MT development, it was remarkable that in contrast to the heated "rationalists vs. empiricists" debate which took place at the same Montreal hotel at TMI in 1992 between groups espousing "rule-based" or "statistical" MT, almost everyone now agrees that the so-called "hybrid" approach, i.e., combining two or more development methodologies, is the most appropriate way to go about building machine translation systems. As evidence of the impact of automated tools and readily available electronic resources on MT development, Dale Bostad of the National Air Intelligence Center noted that the development cycle for a production MT system in a new source language has dropped from 15 years to 5 years.

The conference opened with talks by Muriel Vasconcellos and Mary Flanagan, both of whom focused on the current intense activity on the Internet and other online services. We were given notice by Mary Flanagan of CompuServe that "on-line" is where all the action is, and is going to stay. Remarkable statistics were provided in both presentations, including a count of over 30 sites offering MT on line. Although developers who continue to prosper "off-line" would disagree with her assessment that the only action is on-line, Ms. Flanagan has considerable ammunition to back her claims that the largest volume use of MT has moved on-line. Interestingly, the success of the on-line service has far outstripped CompuServe's hopes or expectations in implementing it. The initiative of offering MT on-line to allow near-real-time translation of email and international dialogues in selected forums has clearly matured into a service that people are using for real communication. Two years ago Mary reported that on-line translation had just begun, but that with the novelty of MT, the discussion tended to drift off of the forum topic to commentary on the MT output instead. Now the World Community Forum, which subscribers pay a surcharge to participate in, together with CompuServe's document translation service is translating 32 million words a year. An excellent use of "crummy" MT (K.W.Church & E.Hovy: Good applications for crummy machine translation, *Machine Translation* 8 (4), 1993, 239-258)!

In a back-to-back session on specialized vs. general-purpose MT, John Chandioux gave a very entertaining account of the implementation of the Meteo MT system for weather report translation for the Atlanta Olympic games. It was a time-compressed example of all the implementation problems that can occur for regular users, and many more that were unique to the Olympic games. All were quickly resolved for a remarkable MT success story. This was followed by Dale Bostad's description of MT use in the U.S. Air Force, and U.S. Government. As in Marjorie León's account of giving access to non-translators via the PAHO LAN, and John Hutchins' description of increased use of MT at the European Commission, Bostad's account underscored the fact that sometimes speed is more important than quality. The keys to widespread use of MT for information in an organization are 1) publicity and training - make sure people know it is available and how to use it, and 2) direct access.

Historically, the established MT developers have had a hard time finding a forum to present their work. Theoretical forums tend to reject papers from commercial developers because the work is not groundbreaking, and yet complaints are still heard that nothing is published and people don't know what commercial developers are doing. Patricia Schmid and Claudia Gdaniec of Logos have been doing their part to remedy this situation, presenting papers on the LOGOS methodology. Their paper which follows the system's evolution across decades and hardware platforms makes a solid contribution to the understanding of the opportunities and constraints of developing production MT systems.

The AMTA by design includes representatives from three segments of the MT community:

developers, researchers, and users. Historically the number of committed users has been relatively small, and we saw some of the same users conference after conference. Over time, however, users are becoming a larger and more diverse group. This time, experienced MT users spoke out in a variety of panels. The input and participation of this group is very enriching. It was impressive to meet a number of users and "shoppers" at the conference. Several large companies had sent representatives to investigate and develop expertise in MT in anticipation of large-scale implementation for localization of product literature.

In a back-to-back session on "MT in Practice", Ron Fournier, who has been using MT in his translation agency Lexitech for 9 years, was positive about his own experience. He was also careful to balance his remarks with cautions on the investment necessary for hardware, termbank development, and training to run a fully automated MT shop effectively. In the same session, Marjorie León presented an impressive, well qualified, account of time and cost savings at PAHO using SPANAM and ENGSPAN. She also provided some practical advice on the conditions for successful implementation of MT.

Evaluation has been a perennial topic of interest and contention. In the panel entitled "What can we learn from users", Edith Westfall proposed that AMTA come up with standards and an "AMTA stamp of approval" which could be displayed on the packages of systems which comply. Also of interest was a back-to-back session-turned-debate on evaluation, in which John White (PRC) took a stand in favor of core MT technology evaluation to promote development of the most promising basic approaches to MT. Ed Hovy (ISI) proposed an alternative application-based evaluation to help users select the best MT system for the particular translation task in which they plan to use MT. Hovy objected to the notion that a "core MT technology" or "core engine" can really be identified given the diversity of approaches now being pursued. (In summary, we still appear to be a long way from a generally accepted standard or method for evaluation.)

Reports from John Hutchins on MT in Europe and Hitoshi Iida on MT in Asia revealed that there is a proliferation of consumer MT and specialized MT products. In Europe there is considerable focus on translator workstations, with limited interest in general purpose MT systems. Inexpensive or very specialized (e.g. patents only) systems are gaining some popularity primarily with non-translators. In Japan, the many systems developed by electronics manufacturers in the 1980s have been joined by a number of less expensive feature-driven consumer systems, including a few which are just for web browsing.

A prediction made in 1990 that commercial MT in the '90s would consist of a very small number of MT systems priced in the several-hundred-thousand-dollar range could hardly be further from what has evolved. The last 6 years have instead produced a larger and better populated market of MT systems with increasing competition in price and features. At last year's Summit, a great romance was blooming in the hearts of developers for users. No developer dared speak about their product without including a statement of devotion to user needs and "environmental" concerns. Ultimately though, the user's best friend is competition. Nothing else has inspired such authentic improvements in products, and that is very healthy for the industry as a whole.

The conference closed with an invited presentation by Yorick Wilks, which explored the possibility of using statistical techniques for sense disambiguation of polysemous words, in the same way that such techniques have been successfully used for part-of-speech disambiguation.

There were a number of opportunities for conference participants to mingle, including a Welcome Reception Tuesday evening, continental breakfasts, luncheons, and other breaks. Thursday evening was enlivened by the AMTA general membership meeting. The newly elected officers were announced, and reports given by the various special interest groups. Friday afternoon walking tours of Montreal gave out-of-towners an opportunity to enjoy local color in the brisk autumn weather with local guides. Friday evening the banquet was held in the Montreal Museum of History and Archeology. This unique event began with cocktails and a tour of the sheltered excavations of the successive settlements in old Montreal.

The following companies exhibited and demonstrated their products in the exhibition room: Machina Sapiens (Montreal, Canada); SYSTRAN Software, Inc. (La Jolla, CA); NeocorTech LLC (San Diego, CA); MCB Systems (San Diego, CA); CITI-TAO (Montreal, Canada); LOGOS

Corporation (Santa Clara, CA); Apptek, Inc. (Mclean, VA); John Chandioux, Experts-Conseils, Inc. (Montreal, Canada); Globalink (San Diego, CA).

Stephen Helmreich (CRL, New Mexico State University)

Mix together an amicable crew of MT producers, MT researchers, and MT consumers together with a cosmopolitan city and congenial hosts and you have all the ingredients for a completely enjoyable and interesting MT conference.

The second AMTA conference was held at the Radisson hotel in downtown Montreal with the theme: Expanding MT Horizons. Pierre Isabelle and the CITI group served as the local hosts, while Elliot Macklovitch was responsible for local arrangements. All involved did an excellent job. Organizer Ed Hovy is to be congratulated for his work in getting the party together and providing the setting that allowed it to be a success. And Muriel Vasconcelles is to be thanked (as she was at the conference) for getting it all started in the first place.

Representatives of most of the major theoretical approaches to MT were present. Although the IBM group was not physically present, their spirit remained alive, particularly in the person and work of Dekai Wu. ISI (in the person of Ed Hvoy and Kevin Knight) provided a tutorial on statistical techniques for MT. All the tutorials I understand were well-attended, often to the extent that there were insufficient copies of notes and handouts available. This may in part have been due to some glitches in the registration procedure. In all, there were 6 tutorials ranging from the aforementioned statistical techniques to criteria for purchasing an MT system.

One suggestion for future conferences. Often a tutorial simply becomes an excuse for researchers to discuss their own current work. While this is often of interest, it might be worthwhile for tutorials to actually provide somewhat broader introductions to topics, rather than more specific and detailed information about the tutor's work.

Preceding the conference was a pre-workshop on Interlinguas, organized by Dr. David Farwell of CRL. There were about 2 dozen participants in what was billed as a workshop to plan a workshop. Each presenter had only 10 minutes and significant time was devoted to discussion around five central topics. A number of issues crystallized which will hopefully serve as the focus for future such workshops. These issues included: the contrast between deep and shallow interlinguas, the question of whether ILs are really a kind of NL, the issue of how much information (particularly language specific information at one end and inferential information at the top end) should be encoded in an Interlingua.

Much of the conference was entirely in dual sessions. This arrangment made it easier for attendees to slip away ("Oh, you didn't see me because I was at the other session."); it also made it much more difficult to hear everything you wanted to. So, to participate fully, you needed not only to be at every session, but to order videotapes of the conference as well. Although this method of presentation allowed for more leisurely talks and discussion, given the relatively narrow focus of the conference and the familiarity of most participants with a broad spectrum within the field, a single venue system might be preferable. The dual presentations also made it difficult to squeeze in time to look at the commercial systems or to attend the demos.

Speaking of demos, it appears that the customs people gave no little difficulty to US demo'ers. And, in one case, UPS managed to smash the glass on a OCR scanner. Those problems and the usual glitches meant that at least three demos had difficulties during their presentations. Maybe some helpful hints to US participants about customs regulations might have been helpful.

I found the panel discussions and invited lectures the most intriguing, and except when an exceptional paper drew my attention, or professional courtesy or direct participation required my attendance at a talk, I generally gravitated towards the large lecture hall.

Several items struck my attention. The big news was the large boom in on-line translation services: cheap and with no guarantee of any quality whatsoever. For some, this appeared a tremendous opportunity, for others it seemed to be perhaps sounding the end of serious large-scale attempts at high-quality machine translation. In either case, it is something that will be an important

part of the MT scene for the next years.

Another interesting proposal was that of Pierre Isabelle for completely user-controlled translation. As users make choices (by typing the beginning or continuation of their own translation), the MT system supplements that translation by completing whenever possible from its store of possible translation variants.

Yorick Wilks suggestion that semantic sense tagging to the homograph level might well be accomplished by the conjoined use of several orthogonal techniques was also exciting. In addition, I was, of course, taken in by Sergei Nirenburg's use of direct quotes from a 1960's Bar Hillel paper to comment on the current state of the art.

The debate between Ed Hovy and John White generated the most heat, though perhaps not the most light. It seems to me that neither John's insistence on a single standard for measuring throughput, nor Ed's desire for a more flexible set of standards for measuring performance were really geared toward evaluation of research systems.

To use their analogy, you can't simply evaluate an engine (John White), nor yet an engine geared toward a specific purpose (engines for dune buggies, sports cars, and trucks), but to get improvements in overall engine performance, you must evaluate independently the performance of various parts: sparkplugs, pistons, fuel injection systems, etc.)

So in research, it should be possible to concentrate on producing a better part-of-speech tagger, or grammar, or analyzer, or proper-name recognizer, without having to worry about its actual use in a performance model.

In summary, the second annual AMTA conference provided a lovely setting for a congenial group of people from all parts of the MT field to meet and exchange ideas, business cards, and slides.

The Proceedings and Videos of the AMTA-96 Conference are available from the AMTA focal point at phone/fax (703) 860-1027 or e-mail AMTAinfo@aol.com.

PRODUCTS and SYSTEMS

Globalink's Web Translator Software To be Bundled With Products by Accent Software

Globalink, Inc. has announced in a recent licensing agreement with Accent Software International, Ltd., that will allow Accent to distribute Globalink's one-of-a-kind Internet translation product, Globalink Web Translator. The initial OEM license will be in the amount of $200,000 for the first year and has the potential to be renewed.

Globalink Web Translator translates Web sites (HTML documents) to and from English into Spanish, French, Italian and German. By contrast, Accent's products provide character recognition for foreign language alphabets so that language proficient users can create documents in different languages.

The agreement is part of Globalink’s continuing effort to expand and diversify product distribution and promotion. It is Accent's intent to bundle Globalink Web Translator with various Accent foreign character support packages to be sold worldwide to 3 target markets; Internet Service Providers; Computer Manufacturers; and Modem Manufacturers.

Alis Technologies Unveils its Global Communication and Translation Suite

Alis Technologies has brought together several companies among the leaders in the language industry to create Columbus, code name for a suite of communication and translation tools and services designed to offer corporations the most comprehensive, convenient and cost-effective

language solution on the market. Berlitz International, Globalink, Logos Corporation and SYSTRAN Software are the first companies to partner with Alis in this ground-breaking project.

Columbus will provide users assistance in displaying, creating, editing and understanding multilingual Internet/Intranet content, e-mail messages and internal documents. It will achieve this through the combined use of Automatic Translation Technology at both client and server levels, and online human translation services. To ensure an integrated approach, Columbus is designed to inter-operate seamlessly with the leading browsers (Microsoft Internet Explorer, Netscape Navigator and Alis Tango), and leading e-mail systems.

In order to answer the pressing communication needs of corporations wanting to be active players on the global scene, Alis has developed the Tango product line, which allows users around the world to browse the Web, exchange e-mail messages and create attractive Web pages in over 90 languages. And now, the Tango family will evolve to a truly complete corporate solution with the addition of the Columbus Global Communication and Translation Suite.

"With the tremendous growth of corporate Intranets and the increasing globalization of markets, companies need to communicate at a global level in order to increase their competitiveness. This means being able to create, access, exchange and understand information in the languages of their customers, employees and partners in a timely and cost-effective manner," said Claude Lemay, President and Chief Executive Officer, Alis Technologies. "Alis' Tango product line, including Columbus, will give corporations an edge in achieving this goal."

Since its incorporation in Montreal in 1981, Alis Technologies has been providing information technology solutions for people who work in languages other than English. Its expertise covers the languages of Asia, Eastern and Western Europe, as well as the Middle East. Recognized internationally as a leader in the field of language technology, Alis is an active member of several committees establishing multilingual standards all over the world. Applying its language-handling expertise to the Internet, Alis has taken steps to facilitate global access to this network of networks by creating a multilingual product line for the Internet called Tango. Alis' head office is located in Montreal, with field offices in Vienna, Dubai, Cairo, Tokyo and Hong Kong, and sales infrastructures in the United States, Mexico, Germany, Switzerland, Italy, Sweden and France.

METAL and its Users

Ursula Bernhard

In May 1995 the German news magazine "Der Spiegel" announced that Siemens Nixdorf Information Systems (SNI) was ending completely its involvement in machine translation technology and therefore all further developments of the METAL translation system. In August 1995 Sietec, the subsidiary of SNI responsible for MT, informed the users of METAL that an investigation had taken place within the framework of the restructuring of Siemens Nixdorf and Sietec and had concluded that systems such as METAL would not be economic products on the world market, although it conceded that machine translation could be continued in the context of client-specific solutions.

METAL users learned at their meeting at the end of September 1995 that the UNIX-based version of METAL would no longer be developed and supported. Non-exclusive rights to METAL had been taken over by, on the one hand, the Office of Foreign Affairs in Munich (a national body subordinate to the Federal Chancellery), and on the one hand, by the Gesellschaft für Multilinguale Systeme (GMS) in Munich and Berlin, which has developed a METAL component for Russian into German. They intend to maintain the METAL technology and provide a service for customers. The meeting was informed that METAL was being converted into C++ and that it was planned to market at Cebit 1996 a PC-based version for end-users in cooperation with a German publisher. In addition there were plans to launch at the Frankfurt Book Fair in autumn 1996 a system intended for the professional translator, also PC-based, and to bring out in early 1997 a so-called high-end version for companies, realised on a client-server configuration. In the meantime, as announced at Cebit 1996, the PC-version has been marketed as Langenscheidt T1, costing DM. 298 [see MTNI#15: ?]

Users of the old UNIX-based METAL system suffered in the meantime an unnerving experience. On the 4th July 1996 (US Independence Day!) METAL systems went down all over the world with the cryptic message: "The scheduler is attempting to recover from error xn..." After hectic email and faxes between METAL users and GMS it emerged that the error lay in the subroutine of another installation. It was a "time-related occurrence" in a release which had long been superseded by a new upgrade. Unfortunately, the more recent upgrades of the subroutine could not be inserted in the METAL software without further work. Fortunately, GMS felt responsible for the old METAL system and delivered a patch allowing METAL to run again. So the old METAL customers, e.g. SAP AG and Boehringer Ingelheim, could carry on. But there was slight unease: when would the old METAL system suffer the next "time-related occurrence" in some other sub-routine?

With this experience the users of METAL looked forward to their meeting on 26th September 1996 with particular interest and expecting to get solid information for the future of their applications. Seven METAL installations were represented at the meeting. It was learnt that although the METAL system was to be further developed and serviced by the Lant company in Belgium, this firm was not contractually committed to be active in Germany. The meeting heard also that Langenscheidt was going to market at the Frankfurt Book Fair in October 1996 the METAL-successor product T1 for the language pairs English-Spanish and Spanish-English (according to more recent information this version will now be available in March 1997) and a product T1 Plus for German-English and English-German with expanded dictionaries (this version in fact appeared on 23 December 1996.) It was also reported that the version for professional translators which had previously been announced for this occasion would not now be available for the language pairs German-English and English-German until Cebit 1997. This version, probably to be called T1 Profi, will include a translation memory, an HTML converter, an interface for terminology importation and an expanded dictionary coding tool. It is to cost DM. 598. In mid 1997 will be launched the T1 Standard version for Russian-German and the T1-Profi version for English-Spanish and Spanish-English. So far 5,000 copies have been sold of the T1 Standard version for German-English and English-German.

The high-end version, now urgently required by the METAL users present at the meeting, will be available towards the end of 1997. In its first form it will, according to GMS statements, not be a client-server configuration but rather a 'frontend-backend' multi-user version including an SGML interface and permitting the transfer of terminology from old METAL systems. Nevertheless, transfer will be possible only with technical support from GMS. Whether there will be a second high-end version for a larger number of users (15 to 20) is questionable, since it depends on the level of interest. In any case, the high-end version is to be platform independent and JAVA oriented.

All METAL users present at the meeting were interested in the high-end version of the METAL successor product, and it was decided to continue the user group for the representation of this interest and for the exchange of information. The secretariat has been taken over by SAG AG in Walldorf.

Whatever Happened to M....?

Lutz Graunitz

Whatever happened to the translation technology formerly known as M....? (*n., adj., & v.-n. 1A* any of a class of chemical elements such as gold, silver, iron and tin, usu. Lustrous ductile solids and good conductors of heat and electricity and forming basic oxides.) - C*oncise Oxford Dictionary, Thumb Index Edition*

Overview

The translation technology formerly known as **M....** was developed by **S....** over a long period of time, approximately 18 years, and at a multi-million dollar cost. Unfortunately this grand investment in a very worthwhile technology has until now not been able to repay S.... with a commensurate sales revenue volume. This high-end translation engine has faced a number of difficulties common to the

MT industry. The premier challenge faced by the MT industry is the disparity between the expectations of users and the providers' ability to fulfil those expectations.

This disparity is no doubt a major stumbling block for those translation technology providers and their potential users who aim to draw up a sound business case for the implementation of MT technology in a for-profit business environment.

Scope

In this short paper we will posit why M.... was mothballed, relate how this technology metamorphosed and explain why bees and Concordes can fly.

In Mothballs

Q.: Why was M.... mothballed?

A.: Money and Core Business

If you, dear reader, want only the synopsis, then please skip forward to Metamorphosis.

The M.... technology has been acknowledged by its peers as one of the best MT technologies in the high-end MT market for those language pairs which it offers. Why then was it ignominiously withdrawn from the market? The reasons for this withdrawal are several and complex:

Marketing and Revenue

The marketing strategy employed by S.... emphasized the sales of M.... as though it were a shrink-wrapped package that one could parachute into any organization that had about $US 100.000 to spare. With the notable exception of SAP AG and a few other M.... customers, attempts to control the quality of the MT input were few and far between. Consequently the results produced ranged from poor to dismal. The general reaction of translators was: "I would have been better off to do the complete translation myself and forget this MT altogether". Obviously, something was awry with S....'s marketing approach. A consequence of this strategy was flagging sales and a cost for maintaining the development team which far exceeded the spotty annual revenue.

Changing Times

In these rapidly changing times (small delta, maximum change), large organizations based on a hierarchical, bureaucratic organizational model have discovered that they lack the flexibility to adapt their business stucture to the paradigm shift to the network model of a global economy. In order to regain flexibility, many companies are slimming down by going on an organizational diet whose mandate is to concentrate only on core business. In the case of S...., sixteenth largest company on the planet, it is so big and so diversified that it does not exactly know what it's core business is. To answer the foregoing question, McK.... consultants were brought in to provide an objective opinion. It appears that this very competent organization is used as a magnifying glass to examine most or all of S....'s business units to determine which ones are core and which not. The group responsible for the M.... technology was among the earliest business units to be examined. It was found wanting. It was disbanded.

Metamorphosis

The metamorphosis of **"The technology formerly known as M...."** into a new, elegant and superior way "to do MT" came about as the result of a serendipitous confluence of events. Just about the time that S.... decided to disband the M.... group, a technology for using the M.... translation engine to produce controlled English from unrestricted English input became available. A general opinion solidified that the complete language technology would yield the best results if the following conditions were met:

1. Market the technology only as a project which incorporates the product.
2. Make controlled English input a pre-requisite for high-volume technical MT.
3. Market only to clients who are willing to change the document authoring and production process to accommodate the requirements of the controlled English and the MT technology.
4. Ensure that the client is willing to invest in the lexicon work required to fully build up a domain-specific dictionary covering all of the client's documentation.
5. Ensure that the client is willing to invest in the controlled English grammar rule

modifications required to tune the system to the client's technical domain and authoring guidelines.

The final factor in this confluence of events was that the first project which met all of the above requirements was won by S.... after a fierce and prolonged competition with the best in the business. As the M.... division was just in the process of being disbanded, this resulted in a quandary for S.... What to do now? The M.... marketing representative in Toronto incorporated the Max Delta Corporation (what else?) with S....'s encouragement, obtained the rights from S.... to continue to market the technology (but don't use the M.... name); found a competent partner to carry out the work; and proceeded to implement the CASL project (Controlled Automotive Service Language) for the General Motors Corporation of Detroit, U.S.A.

Why Bees and Concordes can fly

We used to think that according to the laws of physics bees should not be able to fly. According to the laws of linguistics, MT should not be able to fly. Until now! Imagine this: You take your originally authored English technical text and pass it through the controlled English checker. The checker gives feedback and suggestions as to how to substantially simplify and disambiguate the text and after appropriate iterations you are ready for MT. At this point, there is little need for further pre-editing for the following reason: Since the controlled English input will be analyzed for the MT phase by the exact same parser that has already parsed it in the controlled English phase, therefore we know that we have high-quality MT input and that most sentences will parse.

In short, the recipe utilizes the controlled English checker as a catalyst that raises MT accuracy and productivity to the level where a sound business case can be made for cost savings and time savings through the implementation of MT. Obviously not every business can afford to take advantage of these savings; but those who have a sufficient requirement in a specialized technical domain with a constantly increasing volume of originally authored text will ignore this new development only at their peril. To avoid becoming imperilled, contact: Lutz Graunitz, Marketing Manager, Max Delta Corporation, 7 Helen Avenue, Thornhill, ON, Canada L4J 1J6 (Tel: +1 905 764-3900; Fax: +1 905 764-2701; Email: lutz@maxdelta.com)

MDC's motto: IN SIMPLICITATE VERITAS

Another User for the TranSmart Finnish-English System

[Press release 16.10.1996]
[This was translated from Finnish by the TranSmart MT system]

Raahe Steel brings into use the machine translation system which translates texts from Finnish into English.

Rautaruukki Raahe Steel brings into use in its data network the TranSmart machine translation system developed by the Kielikone Oy which translates texts from Finnish into English. Raahe Steel and Kielikone have developed this system together so that it is able to produce a relatively good draft translation from the technical texts which are typical to the steel works.

The development project was begun in 1994. The target was to accelerate translation and to collect the own lexicon of Rautaruukki which standardises the used terminology and brings it within the reach of everybody. Rautaruukki has got economical support from Tekes to development work.

An essential part of the development work has been the collecting of the lexicon of Raahe Steel and the editing of it to suit a system. There are about 53000 entries at the moment in the basic lexicon of the machine translation system. Moreover there are about 7000 entries in the lexicon of Raahe Steel. The lexicon of Raahe Steel contains the processes of Raahe Steel Works and the products but also the quality terms and the economy terms and bookkeeping terms. The lexicon is in an electronic shape and it serves except in a machine translation system also as a separate lexicon in the word library in the data network of the group.

A significant part of the development work of the machine translation system has been the

testing of the texts which are typical to Raahe Steel and of the collected lexicon and the tuning of the system so that it is able to translate the texts produced by the employees of Rautaruukki. Kielikone has been responsible for this part.

The TranSmart system leans on the grammar of the Finnish language, in other words the program analyses first a Finnish sentence and then will translate it into English. The text to be translated has to be in an electronic shape and a good Finnish. In spite of this the translation produced by the system is always a draft translation and requires checking.

Rautaruukki continues its participation in the development work of the translation system. Kielikone gets a new test material and the TranSmart system is taught to translate the texts of the employees of Rautaruukki still better. The terminology used by the Strip Products and Metform industrial divisions is added to the system

Kielikone has developed the TranSmart machine translation system for more than ten years. First Kielikone got support from Sitra, after that from Tekes and later also from a few industrial companies.

Supplementary information: Superior of the translation office of Raahe Steel Maija Räsänen, telephone number (08)849 4373.

News on PARS systems

Michael S.Blekhman

Lingvistica '93 Co. has released a package of three bi-directional machine-translation systems compatible with WinWord 7.0 and embracing the three languages broadly used in Ukraine:

- PARS (Russian-English-Russian)
- PARS/U (Ukrainian-English-Ukrainian)
- RUMP (Russian-Ukrainian-Russian).

All the systems run under Windows '95 and can be started directly from MS Word 7.0. When the systems are installed on the user's computer, the "Translate" item is added to the main menu of WinWord 7.0, comprising 3 submenus, for PARS, PARS/U, and RUMP, respectively. You may open a file in WinWord 7.0, have it translated, and the target text will be pasted in a separate window, under the source one, marking polysemantic words with asterisks so that the user could select a more appropriate one where necessary. New fords and phrases can be entered to any of the dictionaries directly from WinWord 7.0.

In this way, you can, for example, have a text translated from Ukrainian, which is the state language in Ukraine, into English and Russian, also broadly used in this country, and display all the three of them in three separate windows. For more information contact: Dr.Michael S.Blekhman (E-mail: blekhman@lotus.kharkov.kpi.ua)

Fujitsu ATLAS on Web

Fujitsu announce the availability of the Atlas MT system on their Web site: www.fujitsu.com.

Software Available

MORPHY: Software for German Morphology

Reinhart Rapp

[From LINGUIST List]

Morphy, an integrated tool for German morphology and statistical part-of-speech tagging developed by Wolfgang Lezius (University of Paderborn, cognitive Psychology) is now available on the WWW under the following URL: http://www-psycho.uni-paderborn.de/lezius/lezius.html

The package includes a lexicon of about 100 000 word forms and documentation (in

German). An overview in English is given in the paper "A Morphology System and POS-Tagger for German", which can be downloaded from the same URL. The tool is intended to make some well established methods (e.g. Church's tagging algorithm), which have been adapted to German, widely available. For this reason it is PC-based, and great emphasis has been put on user-friendliness. The system is suitable for linguistic research, but also for teaching purposes in CL classes and for use by second language learners.

Release of GATE Text Engineering Software

[From LINGUIST List]
The first release of Sheffield University's General Architecture for Text Engineering (GATE) is now available. Details can be found at: http://www.dcs.shef.ac.uk/research/groups/nlp/gate

GATE is a software environment which supports researchers in Natural Language Processing (NLP) and Computational Linguistics (CL), and developers who are producing and delivering Language Engineering (LE) systems.

GATE includes an implementation of the system architecture proposed as part of the DARPA TIPSTER Text Program. This aims to provide: a communication and control infrastructure for linking together sets of LE software modules; and plug-and-play module interchangeability allowing users to select from different suppliers standard access modes to common data resources (lexica, corpora, etc). GATE also provides an interface for the addition and operation of TIPSTER compliant software modules, and for the development of systems made up of such modules, including visualisation tools for parse trees, co-reference chains, and POS tags, and support for document collection management. GATE is not specific to any theoretical approach nor to any application area within LE, NLP or CL. However, with the aim of promoting collaborative research on Information Extraction (IE) and of illustrating GATE in action, the initial distribution of the system includes a set of modules which implements a full-scale MUC-6-compatible IE system called VIE (the Vanilla IE system.) VIE contains modules for: tokenisation, sentence-splitting, stemming, POS-tagging, named-entity identification, parsing, discourse interpretation (including co-reference resolution).

To obtain GATE, please retrieve and complete a licence from:

http://www.dcs.shef.ac.uk/research/groups/nlp/gate/software.html

On receipt of your licence we will mail you a password which will enable you to download the system. GATE is free for research purposes to academic organisations and by arrangement for research purposes to commercial organisations. GATE is funded by the U.K. Engineering and Physical Sciences Research Council (EPSRC).

For further information contact: gate@dcs.shef.ac.uk

Conference Reports

Center for Machine Translation
Tenth Year Anniversary Symposium
October 7 and 8, 1996

Winfield Scott Bennett

While the symposium was primarily a celebration of the 10th anniversary of the Center for Machine Translation, it was also an inauguration for the Language Technologies Institute (LTI). The LTI is a new CMU entity headed by Jaime Carbonell, which subsumes the Center for Machine Translation. The symposium consisted entirely of invited presentations and panels. Additionally participants were able to see live demonstrations of a number of the project prototypes on the afternoon of the first day.

Each of the two days began with a presentation by an invited speaker. The first speaker was Prof. Makato Nagao, Kyoto University, who spoke about his work with digital libraries. This was not a signal that our distinguished colleague from Japan is abandoning MT; the searches for such

libraries will include on-the-fly automatic translation. Ultimately the library is to allow a number of functions including bookmarking and note taking, email and Net access from within the library, hypertexting and speech synthesis, as well as MT. Color and video image searches as well as those for text will be part of the functionality. This was a particularly appropriate beginning to the symposium since this work involves a number of sub-fields in NLP.

The invited speaker the second day was Prof. Christian Boitet, Université Joseph Fourier (Grenoble), GETA and CLIPS. His particular focus was on the historical development of the "pivot" approach proposed first by B. Vauquois, the original director of CETA and then GETA, in the 1960's. The pivot approach has been key in the GETA systems throughout its history, but, as Prof. Boitet showed, the ideas of pivots underlie a number of other MT paradigms.

In addition to the presentations which focused on CMU's projects, there were three panels:

* What it takes to bring MT to market
* What hard problems are left to solve in MT
* Language Technologies in the year 2020

The first panel was the most pragmatic of all the presentations and was inspired by CMT's experience in the creation of CATALYST, the customized version of KANT for Caterpillar. Eric Nyberg, CMU, was moderator and gave a presentation on the CATALYST experience; I spoke about preparing for MT markets from a commercial standpoint; Michael McCord, IBM, spoke about the commercial version of the LMT system now marketed in Germany and Muriel Vasconcellos, AMTA, spoke about her view of emerging MT markets. The panel was interesting particularly in the mixture of panelists.

The panel on hard problems was moderated by Lori Levin CMU, with Nobuo Hataoka, Hitachi, Sergei Nirenburg, New Mexico State University, Alex Waibel, CMU, and Yorick Wilks, University of Sheffield, each contributing his or her ideas. The hard problems I found most notable were:

* dealing with handwritten texts
* face-tracking (for extralinguistic interpretation as an aid to speech-to-speech MT)
* speech recognition
* users' acceptance
* low accuracy
* failure of high quality research to become products

I was struck particularly by Sergei NirenburgIs point that some of the so-called complex problems are really a matter of simple problems taken together, leading to his suggestion that perhaps we need to redefine the complex problems.

The panel on technology in 2020 was chaired by Jaime Carbonell, CMU, and included Hitoshi Iida, ATR Interpreting Telecommunications Laboratories, Makato Nagao, Kyoto University, Steve Richardson, Microsoft Corporation, and Hans Uszkoreit, University of Saarbrücken. The presentations were conjectural, of course, but engaging. A central theme was that much more natural language technology, including MT and speech recognition, would be available to the average user, often as added value technology rather than as separate software. Given the amount of publicity for the "thin client" + Internet architecture, it was curious that all the panelists viewed the future in terms of PCs with much more memory and storage.

Various projects of the Language Technology Institute (LTI) were presented throughout the symposium and also demonstrated on the afternoon of the 7th. The work of the LTI is varied and exciting. In addition to KANT, the projects presented and demonstrated were:

* JANUS/ENTHUSIAST: A speech-to-speech translation project in conjunction with Karlsruhe University. JANUS aims for speech-to-speech MT for English/German to German/English/Japanese in its current phase. An even more ambitious JANUS III is underway; this is to allow translation from five source languages to six targets.
* Diplomat: Another speech-to-speech MT project. This is for English to/from Serbo-Croatian in a military domain. The first prototype runs on a laptop; the second version is to be on a "wearable" computer.
* Informedia: A project designed to provide search and retrieval capability for video and

audio recordings. It combines machine vision, speech recognition and NLP in a generally user friendly interface. Pursuit, the Lycos search engine, is the search mechanism.

* ALICE: Actually two language tutoring packages, one for Spanish (Alicia) and another for Japanese (Alice-chan). Basically the software checks the student's syntax and offers corrections.

* Pronunciation tutor: A package that listens to the student's reading of a second language and makes corrections in both phonetics/phonology and intonation.

* LISTEN: A reading coach modeled on teachers' work with students in reading. Students read aloud and the system offers corrections for words which are not pronounced correctly or which are clearly not understood.

Overall the Tenth Year Anniversary Symposium was a stimulating event for the MT community. It is a bit disappointing that no proceedings were published.

Translating and the Computer 18

London, 14-15 November 1996

John Hutchins

The eighteenth of this annual series of conferences took place for the third year in the imposing surroundings of the Institution for Civil Engineering close by the Houses of Parliament and Westminster Abbey. As always, the organisers had a rich menu on offer.

The proceedings began with Muriel Vasconcellos giving an overview of the recent dramatic changes in the use of MT software and services; in particular the rapid growth of on-line translation of email and Web pages and of PC-based systems for occasional texts, where perfection and high quality output is not demanded, is changing the whole scene of MT development and promotion. The need is even greater for users and developers to have channels of communication and to work together for improvements and new products, and in this vein she outlined the activities and publications of IAMT and its regional associations.

The next speaker, Hans Haller (IAI, Saarbrücken), described a concrete example of collaboration: a project at BMW for the development of a technical documentation system with a multilingual interface for checking grammar and style, and for interactive translation. He was followed by Edward Johnson (University of Cambridge), whose lively and humorous account of the development of controlled languages for international communication among European police forces, was a reminder of how crucially important translation can be in daily life.

Chris Pyne (Sprache & Dokumentation) spoke of his experience as an MT system provider in setting up translation services in German companies, illustrating the 'hurdles' which must be overcome if large-scale MT is to be cost-effective and to satisfy real needs. Ian Gordon (Trados Ltd.) argued that translation services should not be considering MT at all; far more appropriate for their needs were the computer-assisted systems incorporating translation memories. Orlagh Neary (Corel) demonstrated their applicability in her review of the requirements of the software localisation industry (now a major user of computer-based translation tools, with its European centre in Dublin), drawing particularly upon the experience of the Corel Corporation with XL8 and Logos and then development of its own Trinity software as an interface to various tools. She agreed that translation memory and terminology management were the principal aids to ensure good quality.

Other speakers on the first day were Bob Clark (Praetorius Ltd.) and Ruslan Mitkov (University of Wolverhampton); a third speaker could not be present (Abdel-Malek Boualem), but his talk on the MtScript text editor developed in the MULTEXT project is included in the proceedings. Bob Clark described the riches to be found in the on-line 'living dictionary' from Logos, an Italian translation company (with no relation to the MT vendor, Logos Corporation); and Ruslan Mitkov spoke about the various ways in which MT systems (mainly cheaper PC-based systems) had been used in teaching foreign languages, and the plans at his own university to use MT an aid in teaching Italian from scratch.

As normal in these conferences, not all talks were concerned with MT or translation tools

directly. Geoffrey Kingscott (Praetorius Ltd.) on the setting up by the Commission of the European Union of a forum, the European Translation Platform, bringing together representatives of the various European associations concerned with translation and terminology. Sigrid Martin (University of Kent) took up the theme of translator training, and Lyn Robinson (University College London) spoke about the ways in which translators might use the possibilities opened up by access to the Internet. Then Donia Scott (University of Brighton) described the DRAFTER prototype software which generates draft procedural texts directly from databases representing conceptual relationships among items and processes of a specific domain.

After lunch on the second day, participants divided into three groups. In one, representatives of Trados, STAR, Logos and Endeavour Ltd. described their products; in another, Bob Clark gave further insights into the riches of the Logos 'living dictionary'; and in a third, there was a round-table discussion of the use of translation systems in training.

Machine translation was the theme of four final speakers. Harri Arnola (Kielikone Ltd.) spoke on TranSmart, a Finnish-English MT system, developed originally for the Nokia company and now available as a commercial product and as a translation service. The base technology of the Kielikone system is considered sufficiently general and flexible to permit the rapid development of other language pairs. As an adjunct to Arnola's presentation, Pertti Nuutila of Nokia Telecommunications described how the Kielikone system was being used for rough translations of internal company texts. These two speakers were followed later in the day by Klaus Schubert (Fachhochschule Flensburg) on a forthcoming project ALSBALT, which will provide a network-based translation service using various MT products for small and medium-sized businesses in the Baltic Sea region and which will seek to develop or adapt systems for languages of the region not presently covered. The last speaker of the conference was Alan Melby (Brigham Young University), who drew on his wide experience with MT and terminology management to put forward his ideas on the way ahead and what systems vendors should be doing to attract more users. Specifically, he proposed the 'consumer labelling' of systems, the development of a universal framework for controlled languages (both for MT and for spelling and style checkers), the standardisation of text-formatting information (based on elements of SGML), a standard format for terminology exchange, and the marking of links to termbases at the time of text creation.

The proceedings of the conference are available from Aslib, The Association for Information Management, Information House, 20-24 Old Street, London EC1V 9AP (Tel: +44 171 253 4488; Fax: +44 171 430 0514; Email: aslib@aslib.co.uk; WWW: http://www. aslib.co.uk)

SLIG'96, Dublin, 14 October 1996

Reinhard Schäler

"Showcase for a Thriving Industry: Localisation and the Web" was the theme for this year's Conference and Annual Meeting of the Software Localisation Interest Group (SLIG), held on 14th October 1996 at the University Industry Centre, UCD, and organized by the Localisation Resources Centre.

Ireland is the recognized world leader in software localisation and, after the U.S., the world's second largest exporter of software, the conference was told by Aidan Stack, Manager of the International Services Division of the Irish Trade Board, who officially opened SLIG'96. During the day, speakers from companies such as Lotus, Informix, Berlitz, ITP, Stream and Nua presented their localisation strategy for the World Wide Web to more than 100 industry representatives. They proved that Irish-based localisation companies are constantly adapting their
procedures and tools for newly emerging software aimed at the global market and are well prepared to deal with the challenges presented by new technology to localisation experts - including the hottest new technology around, the World Wide Web.

The achievements of SLIG in 1996 were highlighted by Helen Wybrants (ISC-Europe) for the Education and Training Group, Sharon O'Brien (ITP) for the Tools Group, Marion Gunn (Everson Gunn Teoranta) for the Terminology Group and P.J. King (Clockworks Multimedia) for the

Multimedia Group. These achievements included the development of a proposal for the certification of training courses for localisation, the exchange of information and proprietary localisation tools between companies, the establishment of an Internet site where terminology for the translation of software has been made freely available by companies such as ITP, Microsoft, Lotus and Corel, and the compilation of guidelines for the localisation of Multimedia applications.

One of the highlights of the past year was the establishment at UCD of the Localisation Resources Centre as the focus point and the research and support centre for the industry. The Centre was established with funds made available by the Irish Government and the European Regional Development Fund (ERDF). The biggest achievement of the Localisation Resources Centre over the past few months was the setting-up of what is believed to be the world's most complete library of machine translation (MT) and computer assisted translation (CAT) systems with the support of MT and CAT developers who licensed software worth around £200,000 to the Centre free-of-charge.

The future of SLIG was discussed by delegates at the first session of the day. They asked the Localisation Resources Centre to express their surprise and dismay at the recent decision by the IDA, Forbairt and the Irish Trade Board not to support a badly needed survey on the industry. This survey had been proposed at an earlier, general SLIG meeting in the presence of representatives from the semi-state bodies who initially supported it. SLIG supported the Localisation Resources Centre's proposal to organize a series of eight meetings over the coming year. Each of these meetings will deal with issues of particular importance to the industry, such as Localisation and Multimedia, Tools Requirements, Education and Training and Terminology Development. These meetings will be co-ordinated by the Localisation Resources Centre for SLIG.

Localisation and enabling tools for the Web was the title of the first session to focus on the main theme of this year's conference. The discussions were based on a presentation by Ian Dunlop, Development Director with Lotus Development Ireland, who outlined Lotus' web localisation strategy.

Impact of the WEb on the localisation process, the title of the first afternoon session, gave localisation service providers the opportunity to exchange views on the impact of the Internet on the industry. Tom Grogan (Managing Director, ITP) discussed the Role of the Internet in the Implementation of a Global Localisation Process, David Murphy (Software Services Manager, Berlitz) introduced Process Changes - Changing the Way we work, and Paul McBride (Business Unit Manager, Stream) presented a Case Study: SimShip of Internet Localisation Project in 17 Languages.

Localisation of Web content was the last session of the day which included presentations from Gerry McGovern (Managing Director, Nua) on Website Design and Good Marketing Practice, and Patrick Durkin (Pre-Sales Consultant, Informix) on Localising Databases for the Internet. Delegates agreed that this year's highly successful SLIG Conference and Annual Meeting which attracted a record number of participants confirmed the group's representative weight and gave reason for optimism looking forward to SLIG '97.

Presentations given during SLIG '96 - as far as they have been made available by their authors - can be downloaded from the Localisation Resources Centre's web site: http:\\LRC.UCD.IE. The address for the Localisation Resources Centre is: Campus Innovation Centre, Roebuck Castle, University College Dublin, Belfield, Dublin 4, Ireland (Tel. +353-1-7067898; Fax. +353-1-2830669)

Evaluation Workshop

Dublin, 15 October 1996

Automatic Translation - How good is it?

Reinhard Schäler

"Practical Criteria for the Evaluation of Translation Tools for Localisation" was the title of a workshop which attracted a record number of more than 50 participants from world-wide operating software developers and publishers, among them Microsoft, Lotus, Symantec, Corel, Stream, Berlitz and International Translation and Publishing (ITP).

Many automatic translation tools have left the research laboratories and have become

commercially available. Yet, for potential users it is often quite difficult to find out about their particular strengths and weaknesses from an independent source. This is why the Localisation Resources Centre at UCD invited the Software Localisation Interest Group (SLIG) and the European Association for Machine Translation (EAMT) to jointly organize this workshop which took place in Newman House on St. Stephen's Green on 15 October 1996. Over the past 12 months, the Localisation Resources Centre has established a Tools Library where subscribers to the Centre can get independent advice and hands-on experience on a wider variety of translation tools. The Centre's library now houses what is believed to be the world's largest accessible collection of translation tools. Among these is the country's only college-based installation of the LOGOS machine translation system.

World experts in machine translation and evaluation discussed with the participants state-of-the-art automatic translation tools, outlined user requirements and presented models for the evaluation of translation tools.

John Hutchins, President of the European Association for Machine Translation (EAMT), started the day with a brief introduction to Machine Translation (MT), Computer Assisted Translation (CAT) and Evaluation. This was followed by a detailed report on state-of the-art translation tools by Adriane Rinsche, co-author of the OVUM report on translation technology and Managing Director of The Language Technology Centre in Surrey, U.K.

During the second morning session (potential) users of translation tools outlined their requirements. 'Don't Forget the User' was the very apt title of Ana Brady's (Symantec) presentation, who compared her tools requirements to that of a DIY enthusiast - when looking for a workbench, she does not like to be told that the screw driver, hammer and saw come in a different kit which is not quite ready yet but will definitely be delivered within the next 3 months. Orlagh Neary (Corel Corporation) outlined her company's evaluation criteria for translation memory systems. Corel is one of the more experienced users of translation tools in the localisation industry. Although they have now decided to use a particular tool in-house, their experience has shown that it is better to leave it to the translators to decide which tool they want to use. Gunnie Jacobson (Microsoft) introduced the idea of a spreadsheet-based expert system - like decision tool which helps project managers to decide whether a particular project should be localised using their preferred tool.

After lunch, Bente Maegaard (Centre for Language Technology, Denmark) presented some European approaches to evaluation. Prof. Maegaard is one of Europe's most prominent experts in the area of MT evaluation and has participated in many European projects, among them TEMAA and EAGLES. These approaches and their own requirements for evaluation were discussed by the participants in small working groups during the afternoon. The results of these discussions were presented in the last session of the day where it was also decided to organize a follow-up workshop early in 1997. This second workshop will concentrate on industry specific case studies and the presentation of the specification and probably the first prototype of an automated high-level evaluation tool currently being developed by the Localisation Resources Centre together with industrial partners.

News from Europe

European Association for Terminology

The new European Association for Terminology (EAFT) was founded on 3 October 1996 at the Southern Denmark Business School in Kolding on the occasion of the international conference "Language and Business Life". The founder members were drawn from European terminology organisations, language companies and individuals and representatives of businesses in other sectors.

The groundwork of the Association was prepared in the report produced by the EU-funded POINTER project ("Proposals for an Operational Infrastructure for Terminology in Europe"), which recommended the establishment of such an organisation.

In broad terms, terminology can be defined as the specialist vocabularies to be found in

individual domains such as law, medicine and computing. It plays an important role in documentation, and in information retrieval and management, as well as in standardisation, translation and lexicography. In other words, terminology is crucial to communication -- not only across borders, but also among native speakers of a language.

The main goals of the new, non-profit Association are the promotion of plurilingualism and the creation of the co-operative structures needed at European level to further terminology work and training. In addition, the Association aims to heighten awareness of the importance of terminology, especially among decision makers and domain specialists, who use terminology every day without always being aware of the need for clarity and coherence. European level special interest groups will help the exchange of terminological information related to specific domains, as well as promoting and coordinating user-oriented projects and the reusability of existing resources.

Information: Annelise Grinsted (president EAFT), Engstein 1, DK-6000 Kolding, Denmark (Tel: +79 31 11 11; Fax: +79 32 14 48; Email: annelise@ko.hhs.dk)

European Language Resources Association

ELRA (European Language Resources Association), was founded in Luxembourg in February 1995, as a non-profit organization, with the goal of promoting the creation, verification and distribution of Language Resources (LR). In addition to helping users and developers, government agencies, and other interested parties exploit language resources for a wide variety of use, ELRA serves as the European repository for EU-funded language resources, and interact with similar bodies in other parts of the world. Funded in the medium term by membership fees, grants from the European Commission and national governments, together with projects income, the Association will be financially self-supporting in the long run. A 12-member Board is in charge of the strategy and objectives to be adopted by the Association.

Relevant LR are Spoken Databases, Lexica, Grammars, Written Corpora and Terminological Data. These are required for the development of speech and text processing systems for a large number of applications in various information technology areas.

ELRA licences LR for R&D, as soon as these are made available. According to the agreements made between ELRA and its providers, resources can be either reserved for research purposes or used for the development of products and services. Licence agreements, drawn with the support of lawyers, are used for negotiations. Such contracts are available at ELRA office. Lexica, Corpora (spoken and written), Grammars and Tools are described in easy-to-understand presentation forms. All these documents can be found on the ELRA Web site.

ELRA Newsletters

A Newsletter is released by ELRA every Quarter. The next three issues will primarily and alternatively be devoted to each of the 3 Colleges of the Association : spoken, written and terminological resources. The October Newletter (vol.1 n.3) is already available at ELRA office. It is entirely dedicated to Terminology. The next two issues will be dedicated to written and spoken resources.

The ELRA Web site

The ELRA Web site is now available in English and French. The URL addresses are: (English version) http://www.icp.grenet.fr/ELRA/home.html, (French version) http://www.icp. grenet.fr/ELRA/fr/home.html

ELRA Language Resources catalogue

A catalogue of the LR negotiated or under negotiation by ELRA can be found on our Web site. For quotation, please refer to this site or directly to ELRA office. A published version is released twice a year (together with number 1 and 3 of the Newsletter).

Resources appearing in the catalogue are separated according to the three Colleges : spoken

resources, written resources (corpora, lexica and tools) and terminological resources. The catalogue consists of :

1) Spoken resources : 30 databases (recordings from microphone, telephone, continuous speech, isolated words, several languages, etc.).
2) Written resources :
   * 9 monolingual and multilingual corpora
   * 12 monolingual lexica
   * Over 30 multilingual lexical
   * A linguistic software platform and Grammars Development platform
3) Terminological resources : over 90 databases with a wide range of domains and several languages (French, English, German, Spanish, Danish, Italian, Catalan, Turkish, Polish, Portuguese).

ELRA Membership

ELRA members are entitled to substantial discounts on public prices of the resources and other products (such as the Guide for Terminology Agreements and several commercial reports).

ELRA membership is open to any organization, public or private, with full membership (including voting privileges) being available to organizations registered in Europe. Purely for organizational purposes, members will be assigned to one of the Colleges on the basis of their main area of interest. The annual membership fee has been set at a modest ECU 1,000 to encourage broad participation. You may also opt to join more than one College, in which case you will be required to pay multiple membership fees.

Please ask for a membership form or download it from the ELRA Web site. For further information: ELRA/ELDA, 87 Avenue d'Italie, FR-75013 PARIS, FRANCE (Tel: +33 01 45 86 53 00; Fax: +33 01 45 86 44 88; E-mail: elra@calva.net; http://www.icp.grenet.fr/ELRA/ home.html

INTERVAL Project

The Web pages for the INterlinguistic TERminology VALidataion project can be viewed at: http://www.mcs.surrey.ac.uk/interval

Interval is a European terminological resource validation project, co-financed by the EC (DG XIII) within the framework of the Language Engineering programme, which addresses the following issues:

* Multilingual Terminology Resource validation
* Validation methodologies and software tool-kit
* World Wide Web dissemination
* CD-ROM data banks

Language Engineering in Telematics Applications

The European Commission has announced another Call for proposals in the Language Engineering sector of the Telematics Applications Programme 1994-1998. On 17 December 1996 details appeared in the EC Official Journal, and from that data information packages are available with application forms and other related documents from:

DG XIII-E-5 LE Office
Bâtiment Euroforum (0-177),
rue Alcide de Gasperi,
L-2920 Luxembourg
Tel: +352 4301 32886
Fax: +352 4301 34999

The Call-related LE documents are published on the I*M EUROPE Web server (URL: http://www2.echo.lu/langeng/en/lehome.html).

The Call invites project proposals aimed at establishing the feasibility of collaborative RTD scenarios, as well as proposals targeting full-blown projects intended to encompass all relevant phases of work, up to and including demonstration and exploitation plans. The total budget tentatively earmarked for cost-shared RTD projects and accompanying measures established within this Call is in the region of 21 MECU. The evaluation of the submissions will take place in the second quarter of 1997, and successful consortia are expected to undertake their work in the last quarter of 1997, or in the first quarter of 1998.

Preference will be given to proposals aimed at improving, configuring and integrating generic language technologies and resources, and those involving new programme entrants.

Areas of interest include, but are not limited to the following:

*- Internet and WWW based information and transaction services
*- Globalisation of multimedia products and services, and their re-packaging and localisation for new linguistic and cultural markets
*- Foreign language learning, including embedded multimedia language resources for educational and industrial training
*- Telebusiness and electronic commerce
*- Reusable language resources

For complete details and a copy of the Language Engineering project submission guidelines and proposer's Information Package, please visit the Commission's URL <http://www.echo.lu/telematics/> .

Computer-assisted translation at WHO Headquarters

[Press release]

Within the Office of Language Services, the Computer-assisted Translation and Terminology Unit (CTT) is assessing technical options for using computer-assisted translation systems based on "translation memory". With such systems, translators have immediate access to previous translations of portions of the text before them. These reminiscences of previous translations can be accepted, rejected or modified, and the final choice is added to the memory, thus enriching it for future reference. By archiving daily output, the translator would soon have access to an enormous "memory" of ready-made solutions for a considerable number of translation problems.

Several projects are currently under way at the Computer-assisted Translation and Terminology Unit in such areas as electronic document archiving and retrieval; bilingual/multilingual text alignment; computer-assisted translation; translation memory and terminology database management; and speech recognition.

Further information about computer-assisted translation in WHO is available from: Computer-assisted Translation and Terminology Unit (CTT), Office of Language Services, Division of Publishing, Language, and Library Services, World Health Organization (WHO) Headquarters, CH-1211 Geneva 27, Switzerland (Tel: +41 22 791 2317/2458; Fax: +41 22 791 0746; Email: pasteuro@who.ch)

From the Archives...

Fifty years of the computer and translation

John Hutchins

In March this year we mark the fiftieth anniversary of the 'birth' of machine translation, or - more precisely, the beginnings of discussion about how electronic computers might be applied to the task of translating natural languages.

There had, of course, been earlier proposals for the mechanisation of some aspect of the translation process, primarily devices for a mechanical dictionary which would substitute words of one language by words of another. Patents for such devices were issued in 1933 in France to Georges

Artsrouni and in Russia to Petr Troyanskii (Hutchins 1993). In conception, it is possible to trace back the idea of mechanical dictionaries to the seventeenth century, but the mechanisms for implementation did not become reality until the twentieth.

However, these earlier suggestions were unknown to Weaver, and indeed to many other MT pioneers until the late 1950s. In any case, Weaver's inspiration was the potential of the newly developed US computers, whose awesome power gave them the popular name at the time of 'electronic brains'.

It was on 4th March 1947 that Warren Weaver, Director of the Natural Sciences Division of the Rockefeller Foundation, wrote to his mathematician friend Norbert Wiener, shortly to become famous for his writings on cybernetics. Weaver had met Wiener during the War when both were involved in military research: Weaver on ballistics, Wiener on radar and prediction theory. Weaver (1947) wrote as follows:

> One thing I wanted to ask you about is this. A most serious problem, for UNESCO and for the constructive and peaceful future of the planet, is the problem of translation, as it unavoidably affects the communication between peoples. Huxley has recently told me that they are appalled by the magnitude and the importance of the translation job.
>
> Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while.
>
> Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography - methods which I believe succeed even when one does not know what language has been coded - one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."
>
> Have you ever thought about this? As a linguist and expert on computers, do you think it is worth thinking about?

The letter was reproduced by Weaver in his memorandum of July 1949, which effectively launched MT research in the United States -- and the memorandum itself was included in the Locke and Booth collection (Weaver 1955).

Weaver's reference to the possible analogy of cryptography was fully explicable at the time. He had himself heard of an impressive feat of decipherment from Professor Prager at Brown University. It involved a message encoded in Turkish, which was taken to a mathematician ignorant of the language of the original. The text was decoded based on letter frequencies in English, which fortunately matched those of Turkish (after simplification of letters with diacritic marks not found in English). To Weaver, the feat seemed to demonstrate that there were statistical uniformities in all languages that could be used in some way in translation. Weaver was himself a prominent mathematician and was familiar with the work of Claude Shannon on cryptography; later, Weaver collaborated with Shannon on the statistical theory of communication (Shannon and Weaver 1949).

The idea of using computers for translation had evidently occurred to Weaver some time before writing to Wiener. In his autobiography (Weaver 1970) he states: "Early in 1947, having pondered the matter for nearly two years, I started to formulate some ideas about using computers to translate..." Quite possibly he spoke about it to others before writing to Wiener. Bar-Hillel (1952) says that "as early as 1945... Dr.Warren Weaver... started thinking and talking about the possibility..." One of those he may have mentioned it to was Professor J.Desmond Bernal of Birkbeck College (University of London), a physicist much interested in numerical computation, who arranged for Andrew D.Booth to visit the United States to investigate developments in computing.

Booth's first visit was in June 1946. He met Weaver at the Rockefeller Foundation to discuss

the possibility of the Foundation's funding a computer at London University. He then went to the Moore School in Pennsylvania and Princeton University to see and learn about progress on the ENIAC and EDVAC computers. In the following year, he returned for a three month study period at Princeton, funded by a grant from the Rockefeller Foundation. He met Weaver again on 6th March 1947, just two days after Weaver's letter to Wiener. It was on this occasion that Weaver mentioned to Booth the possibility of using the London computer for "non-numerical" applications -- which were more likely to attract American funding support -- and suggested in particular mechanical translation. Booth claimed later in his well-known 'historical introduction' to the collection he edited with William Locke (Booth and Locke 1955) that MT had been discussed by Weaver and himself during the 1946 meeting. However, there is no documentary evidence to support the assertion. For example, in the report of his 1946 visit which Booth submitted to the Rockefeller Foundation, he makes no reference to such an application (Booth 1946), whereas, by contrast, after his 1947 visit, his report to the Foundation in February 1948 includes the following passage: A concluding example, of possible application of the electronic computer, is that of translating from one language into another. We have considered this problem in some detail and it transpires that a machine of the type envisaged could perform this function without any modification in its design. (Booth 1948)
In fact, in a number of articles on MT, Booth states explicitly that the discussion with Weaver took place in March 1947. It is unfortunate that Booth's historical account in the Locke/Booth collection of early MT articles has led so many later writers into error on this point.

There is, in short, no doubt. Although Weaver may have spoken to Bernal and others earlier about the possibility of using computers to translate, the first corroborated and definite mention was made by Warren Weaver in his letter to Norbert Wiener and in his discussion with Andrew Booth, in the early days of March 1947.

References

Bar-Hillel, Y. (1952): Mechanical translation: needs and possibilities. (Presented at MIT conference, June 1952)

Booth, A.D. (1946): Report on visit to American laboratories. 3rd July 1946. (Rockefeller Foundation Archives)

Booth, A.D. (1948): Proposals for the establishment of a centre for the construction and use of an electronic computer, February 12, 1948. (Rockefeller Foundation Archives)

Booth, A.D. and Locke, W.N. (1955): Historical introduction. In: Locke and Booth (1955), 1-14.

Hutchins, J. (1993) The first MT patents, *MT News International* 5, 14-16.

Locke, W.N. and Booth, A.D. (1955), eds.: *Machine translation of languages: fourteen essays*. Cambridge, Mass.: M.I.T.Press.

Shannon, C. and Weaver, W. (1949): *The mathematical theory of communication*. Urbana: University of Illinois Press.

Weaver, W. (1947): Letter to Norbert Wiener, 4 March 1947.

Weaver, W. (1955): Translation. In: Locke and Booth (1955), 15-23.

Weaver, W. (1972): *Scene of change: a lifetime in American science*. New York: Scribner.

CORPORA and SERVICES

ALLEX

[From LINGUIST list]

A group consisting of members from Gothenburg (Sweden), Oslo (Norway) and Harare (Zimbabwe) have been working on a corpus-based lexicon of Shona (monolingual). The whole project goes under the name ALLEX (African Languages Lexicon Project). The first Shona corpus is being converted to TEILite format. It consists mostly of spoken material (interviews). It will be expanded and more printed material will be added in the next phase. The first phase was successfully concluded with the

publication of the dictionary very recently. The next phase will also take up Ndebele, closely related to Zulu. The project leader is Herbert Chimhundu at the University of Zimbabwe, Dept. of African Languages and Literature (Email: allex@zimbix.uz.zw)

LOB Corpus

[From LINGUIST list 7-1666]

The Corpus is a joint enterprise of Lancaster, Oslo, Bergen (the LOB-Corpus). It was compiled at the Norwegian Computing Centre for the Humanities in Bergen; it is a 30-mill. word archive of the English published in 1961. There is a (slightly expensive) CD-ROM available that also includes LOB's American equivalent, the BROWN-Corpus (Univ.,Providence, Rhode Island, USA; also known as FRANCIS/KUCHERA-Corpus),and the LONDON-LUND-Corpus. Information about the CD-ROM can be acquired from the International Computer Archive of Modern English in Oslo: ICAME (e-mail: icame@hd.uib.no) The University of Freiburg, Germany, is presently working on a corpus which will be an update (1990s) of LOB and BROWN.

Latest from Linguistic Data Consortium

RM1: The Resource Management-Word Data Continuous Speech Database. Isolated and Spelled Word Data.

This CD-ROM contains previously-unreleased isolated-word and spell-mode (spelled out words) speech data from the (D)ARPA Resource Management (RM1) Corpus. This data is based on a 600-word subset of the 991-word RM1 vocabulary and contains spoken and spelled words pertaining to the RM1 naval resource management task. This corpus was collected simultaneously as part of the RM1 Continuous Speech Corpus (NIST Speech Discs 2-1-2-4) and contains speech from the same sets of subjects used in RMI.

VAHA: Voice Across Hispanic America

Voice Across Hispanic America (VAHA) is a corpus of Spanish telephone speech, recorded digitally from 915 native speakers of Spanish in various parts of the United States. With nearly 39,000 recorded and transcribed utterances, VAHA will be useful for a variety of research studies, but it is intended primarily for speech technology research and development in telecommunications applications. It is patterned after MACROPHONE (LDC94S21), an American English corpus that is widely used for this purpose.

European Language Newspaper Text

This corpus includes roughly 100 million words of French, 90 million words of German and 15 million words of Portuguese. The European Language Newspaper Text corpus is composed of news text from Associated Press Worldstream, Agence France Presse, Deutsche Presse Agentur and Le Monde.

Further information about the LDC and its available corpora can be accessed on the Linguistic Data Consortium WWW Home Page at URL http://www.ldc.upenn.edu/.

Speech Corpora from CSLU

The Center for Spoken Language Understanding at the Oregon Graduate Institute of Science and Technology is releasing two new telephone speech corpora: 22 Language and Alphadigit. More information available at: http://www.cse.ogi.edu/CSLU/corpora/. As always, the CSLU corpora are available at no cost to universities and other not-for-profit organizations. Companies may obtain speech corpora and other benefits through membership in CSLU's industrial affiliates program (see http://www.cse.ogi.edu/CSLU/memberships/memberships.html).

UCLA short course on "Machine Translation"

On April 2-4 1997, UCLA Extension is presenting again their short course, "Machine Translation", on the UCLA campus in Los Angeles. The tutors are Ed Hovy and Kevin Knight of USC Information Sciences Institute. Full details of the content and tutors will be found in MTNI#13 (February 1996), page 9. The fee remains $1195, which includes the extensive course materials. For additional information and a complete course description, please contact Marcus Hennessy at: Tel: (310) 825-1047; Fax: (310) 206-2815; Email: mhenness@unex.ucla.edu; WWW: http://www.unex.ucla.edu/shortcourses.

PUBLICATIONS ANNOUNCED

Natural Language Engineering

Special issue on Extended finite state models of language. Guest editor Andras Kornai.

In spite of the wide availability of more powerful (context-free, mildly context-sensitive, and even Turing-equivalent) formalisms, the bulk of the applied work on language and sublanguage modeling, especially for the purposes of recognition and topic search, is still performed by various finite state methods. This special issue, based on the proceedings of the W1 workshop at ECAI'96, is open to submissions on finite state methods to text analysis, speech/OCR language modeling, and related CL and NLP tasks, as well as to papers analyzing and possibly extending the domain of finite-state algorithms.

Computational Linguistics

Special Issue on Natural Language Generation. Guest Editors: Robert Dale, Barbara Di Eugenio, Donia Scott.

The automatic generation of natural language texts is an important aspect of most natural language applications, e.g.: systems aimed at achieving interactive dialogue, report or instruction generation, and machine translation. However, Natural Language Generation (NLG) has for a long time been overshadowed by the study of natural language understanding, encompassing tasks such as parsing and interpretation.

The goal of this special issue of Computational Linguistics on Natural Language Generation is to bring together a collection of papers that will attest to the progress of the field and disseminate it to a wider audience. We expect the papers in the Special Issue to address a broad spectrum of issues in NLG, including discourse planning; sentence planning; linguistic realisation; the development of lexical and grammatical resources for generation systems; multilingual generation; multimodal generation; and evaluation issues.

The Language Engineering Directory

Now available. Compiled by Language & Technology S.L. 1996, 398 pages, paperback (Price: US$ 120 / GBP 77.50 / DM 185 / Ptas 15,500 ISBN: 84-922059-0-3)

The Language Engineering Directory is a compendium of organisations, products and services in the language industries and language engineering sectors. The Directory contains information on more than 700 organisations which play a role in shaping the language industries, providing an exciting forum in which researchers, developers, standardisation bodies, professionals, academics, private companies, and users can discuss and provide solutions to the challenges posed by multilingualism in business, academia and society in general.

The Language Engineering Directory will be of interest to a broad spectrum of language professionals including translators, interpreters, terminologists, teachers, computational linguistics, researchers, software developers, consultants and other professionals working in public and private

organisations, and indeed to anyone interested in language technology and language engineering from a technological or commercial perspective.

The Language Engineering Directory has been compiled from questionnaires and telephone interviews commissioned by Directorate-General XIII/E of the Commission of the European Communities in Luxembourg. It is published by Language & Technology S.L., PO Box 2572, 28080 Madrid, Spain. (Fax: (34 1) 519 91 98, Email: 100575.2522@compuserve.com)

Textbook reprinted

Academic Press has announced the reprint of *An introduction to machine translation*, by John Hutchins and Harold Somers (ISBN: 0-12-362830-X). It is available from booksellers at £24.95.

Survey of Research in Japan

The US Department of Commerce has published a report entitled *Survey of Research in Japan*, which Osamu Fujimura of The Ohio State University authored for us, based on site visits he made to Japanese labs. The report is available via the web at: http://www.ta.doc.gov/ aptp/japan/techlit/speech2.htm.

Hard copy available from: Patti O'Neill-Brown, Japan Technology Analyst, Asia Pacific Technology Program Rm. H4410, U.S. Department of Commerce, 14th & Constitution Ave NW, Washington, DC 20230 (Tel: 202-482-6805; Fax: 202-219-3310; Email: PONeillBrown@ doc.gov; WWW: http://www.ta.doc.gov/aptp/aptp.htm)

International Journal of Corpus Linguistics

IJCL presents a wide range of views on the role of corpus linguistics in language research, lexicography and natural language processing (NLP). IJCL seeks to publish research that views language as a social phenomenon that can be investigated empirically on the basis of authentic spoken and written texts. Corpus linguistics specifies corpus design in respect to research interests, provides computational methods of extracting linguistic knowledge, and conceives tools to validate the accuracy of linguistic description. IJCL aims to conciliate the expectations of language industry with the goals of academic research. Corpora are the basic resources in language engineering. It is the linguistic knowledge extracted from corpora that determines the perfomrance on any NLP application. IJCL is a forum to exchange and share expertise, visions as well as information on resources and tools.

For information on contributions and guidelines contact the editor: Dr. Wolfgang Teubert, Institut für deutsche Sprache, Postfach 10 16 21, D-68016 Mannheim (Fax: +49 621 1581 415; Email: IJCL@ids-mannheim.de)

Grammars

Announcement and Call for Papers

GRAMMARS is an international forum for the dissemination of high-level original research in the intersection between mathematical/computational linguistics and formal language theory.

The journal is based on the fact that a certain gap exists between linguists (trying to find adequate formal tools for natural language description) and computer scientists (often unworried about the descriptive applicability of the generative devices they design). Not all linguists are acquainted with the new developments in pure formal language theory, as well as not all computer scientists know the concrete problems linguists deal with and the real needs they have. The aim of the journal is to bring both communities together and to provide a platform for discussion.

Examples of papers fitting the scope of the journal are: those developing new generative devices potentially useful for natural language description, and those clarifying empirical linguistic facts in such a way that they become acceptable for formal language treatment.

We are planning to publish 3 issues, appr. 300 pages, yearly, in March, July and November. The first issue will be available in March, 1997. The journal will be published by the Rovira i Virgili University (Tarragona, Spain). Editor: Carlos Martin-Vide, Research Group in Mathematical Linguistics and Language Engineering (GRLMC), Rovira i Virgili University, Pl. Imperial Tarraco 1, 43005 Tarragona, Spain (Tel.: 34-(9)77-559543; Fax: 34-(9)77-559597;
Email: cmv@astor.urv.es, cmv@tinet.fut.es)

Subscription prices for 1997 (including postage and handling): 3 issues: institutional, USA $ 102; personal, USA $ 51; 1 issue: institutional, USA $ 46; personal, USA $ 23

Please, send your orders to the editor.

EAGLES Results Now Available

[From LINGUIST list 7-1655]
The Expert Advisory Group on Language Engineering Standards (EAGLES) is pleased to announce that results of the EAGLES initiative are now available on the EAGLES WWW server at:

http://www.ilc.pi.cnr.it/EAGLES/home.html

EAGLES (LRE-61-100) is an initiative of the European Commission which aims to accelerate the provision of standards for:

* Very large-scale language resources (such as text corpora, computational lexicons and speech corpora);
* Means of manipulating such knowledge, via computational linguistic formalisms, mark up languages and various software tools;
* Means of assessing and evaluating resources, tools and products.

Numerous well-known companies, research centres, universities and professional bodies across the European Union are collaborating under the aegis of EC DGXIII to produce the EAGLES Guidelines which set out recommendations for de facto standards and for good practice in the above areas of language engineering. Some 200 individuals are involved.

The EAGLES initiative is coordinated by Consorzio Pisa Ricerche, Pisa, Italy which also manages the EAGLES WWW server.

The work towards guidelines was carried out by five Working Groups: Text Corpora; Computational Lexicons; Grammar Formalisms; Evaluation; Spoken Language

As may be appreciated, we are engaged in a massive collaborative exercise that is producing large volumes of documentation. The project is steadily working towards full publishable quality of the documentation and also widespread dissemination of hyperdocuments via the Web. What is made available on the EAGLES WWW server is to be understood in this light. The reader is advised to read the Editors' Introduction which gives information on the initiative in general and describes the currently available documentation. The reader is further requested to note the relative status of documents and their date of revision. This latter information is to be found by following the link 'Browse the EAGLES Guidelines' from the home page.

A major concern is to receive feedback from the wider community on the contents of the EAGLES Guidelines, thus the reader is actively encouraged to contribute through constructive criticism. Details of the feedback mechanism are also available from the home page.

Computational Linguistics and Chinese Language Processing

Contents of vol.1 no.1, 1996:

A survey on automatic speech recognition with an illustrative example on continuous speech recognition of Mandarin (Chin-Hui Lee, Biing-Hwang Juang); Issues in text-to-speech conversion for Mandarin (Chilin Shih, Richard Sproat); A Mandarin text-to-speech system (Sin-Horng Chen, Shaw-Hwa Hwang, Yih-Ru Wang); An overview of corpus-based statistics-oriented (CBSO) techniques for natural language processing (Keh-Yih Su, Tung-Hui Chiang, Jing-Shin Chang); A hybrid approach to machine translation system design (Kuang-Hua Chen, Hsin-Hsi Chen); A model

for robust Chinese parser (Keh-Jiann Chen); Important issues on Chinese information retrieval (Lee-Feng Chien, Hsiao-Tieh Pu).

Membership & Subscriptions:

* Individual Membership in the society (US$50/yr, or NT$1,000/yr) includes a subscription to CLCLP and the Newsletter.
* Institutional Membership (US$1,000/yr, or NT$20,000/yr) includes a subscription to all ROCLING publications including CLCLP, Newsletter, and Conference Proceedings.
* Non-member may order copies of CLCLP at US$100 or NT$2,500 per volume for all destination.

Guidelines for Submissions

Computational Linguistics and Chinese Language Processing invites submission of original research papers in the area of computational linguistics in general and Chinese (natural) language processing in particular. Contribution can be written either in English or Chinese. English will be the primary language of this journal for its international readership. A 600-word extended abstract in English is required for all Chinese contributions. Manuscripts submitted must be previously unpublished and not be under consideration elsewhere. All manuscripts must be submitted in triplicate to: Computational Linguistics and Chinese Language Processing, Professor Keh-jiann Chen, Institute of Information Science, Academia Sinica, Nangkang, Taipei 115, Taiwan

The Balancing Act: Combining Symbolic and Statistical Approaches to Language.

Edited by Judith Klavans and Philip Resnik (Language, Speech, and Communication series.) A Bradford Book, MIT Press, November 1996. 140 pp. ISBN: 0-262-61122-8. $17.50 paper

Symbolic and statistical approaches to language have historically been at odds - the former viewed as difficult to test and therefore perhaps impossible to define, and the latter as descriptive but possibly inadequate. At the heart of the debate are fundamental questions concerning the nature of language, the role of data in building a model or theory, and the impact of the competence-performance distinction on the field of computational linguistics. Currently, there is an increasing realization in both camps that the two approaches have something to offer in achieving common goals.

The eight contributions in this book explore the inevitable "balancing act" that must take place when symbolic and statistical approaches are brought together - including basic choices about what knowledge will be represented symbolically and how it will be obtained, what assumptions underlie the statistical model, what principles motivate the symbolic model, and what the researcher gains by combining approaches.

The topics covered include an examination of the relationship between traditional linguistics and statistical methods, qualitative and quantitative methods of speech translation, study and implementation of combined techniques for automatic extraction of terminology, comparative analysis of the contributions of linguistic cues to a statistical word grouping system, automatic construction of a symbolic parser via statistical techniques, combining linguistic with statistical methods in automatic speech understanding, exploring the nature of transformation-based learning, and a hybrid symbolic/statistical approach to recovering from parser failures.

Recent Advances in Parsing Technology

Edited by Harry Bunt and Masaru Tomita

Parsing technologies are concerned with the automatic decomposition of complex structures into their constituent parts, with structures in formal or natural languages as their main, but certainly not their only, domain of application. The focus of 'Recent Advances in Parsing Technology' is on

parsing technologies for linguistic structures, but it also contains chapters concerned with parsing two or more dimensional languages. It presents an overview of recent developments in this area with an emphasis on new approaches for parsing modern, constraint-based formalisms on stochastic approaches to parsing, and on aspects of integrating syntactic parsing in further processing.

Order information: Recent Advances in Parsing Technology, edited by Harry Bunt and Masaru Tomita (Text, Speech and Language Technology Series, Vol.1) Kluwer Academic Publishers, Dordrecht, August 1996, Hardbound, 432 pp. (ISBN 0-7923-4152-X) NLG 195.00; US$128.00; GB£86.00

Computers and the Humanities

The third number of Volume 30 (1996) of Computers and the Humanities (CHum) has just been published by Kluwer Academic Press.

This issue introduces a new feature of the journal entitled "Debates in Humanities Computing". This first debate in the series treats the controversial topic of statistical methods for authorship attribution, which has recently received unprecedented coverage in the international press: first, concerning the controversy over Richard Abrams' and Donald Foster's assertion of Shakespearean authorship of an obscure elegy, and later (and even more spectacularly), concerning Foster's subsequent attempt to identify the author of "Primary Colors" (Random House, 1996).

The debate presented in this number of Computers and the Humanities includes an attack by Elliot and Valenza on statistical methods used in Shakepearean authorship studies, and Donald Foster's detailed rebuttal of their claims. The regular articles in the issue also report on results of computer-assisted stylistic studies.

The articles in this number of CHum are sure to fuel the continued debate over statistical methods, and is of interest to all those involved in authorship and stylistic studies as well as statistical methods for language analysis generally.

PUBLICATIONS RECEIVED

Journals

AAMT Journal *no.17, December 1996* In Japanese only.

ALPNET news *issue 1, 1997*

Elsnews *vol.5 no.3 (September 1996)*. Contents include: Localisation and language engineering - where's the link? (Rose Lockwood).

Language International *vol.8 no.5 (October 1996).* Contents include: Cultural assumptions in software (Kenneth Keniston). --Upstream of translation: document generation; *vol.8 no.6 (December 1996).* Contents include: Localisation in Norway and Scandinavia (Fergus Geraghty). -- Torte and tautology (Gary Goldfinch) [on TKE 96 conference, Vienna].

LISA Forum Newsletter *vol.5 no.4 (December 1996)*. Contents include: The coming age of teletranslation (Minako O'Hagan). --The cost-effectiveness of today's localisation business (David Brooks). -- Lessons from LA (Robin Bonthrone) [on LISA Forum USA]. -- Multimedia localisation (David MacLeod)

Literary and Linguistic Computing *vol.11 no.4 (November 1996)* Contents include: Virginia Woolf's *The Waves* in French and German waters: a computer assisted study in literary translation (Jan-Mirko Maczewski). -- A hybrid disambiguation model for prepositional phrase attachment

(Haodong Wu and Teiji Furugori).

Machine Translation *vol.11 nos.1-3 (1996), Special issue: Lexical choice*, edited by Leo Wanner. Contents: Lexical choice in text generation and machine translation (Leo Wanner). -- Multilingual generation: the role of telicity in lexical choice and syntactic realization (Bonnie J.Dorr and Mari Broman Olsen). -- Lexical paraphrases in multilingual sentence generation (Manfred Stede). -- On expressing metonymic relations in multiple languages (Helmut Horacek). -- Lexical choice for complex noun phrases: structure, modifiers, and determiners (Michael Elhadad). --Systematic alternatives in lexicalization: the case of gerund translation (Stephan Mehl).

Machine Translation Review: the periodical of the Natural Language Translation Specialist Group of the British Computer Society, *issue no.4 (October 1996)* Contents include: Anaphora and machine translation (Ruslan Mitkov). -- The Power Translator: an evaluation of a PC-based MT system (Derek Lewis)

Multilingual Computing *vol.7 no.2 (1996).* Contents include: Translation services go online! (Edith Westfall). -- Europe's language technology push (Colin Brace)

Software Localisation: quarterly newsletter of the Localisation Resources Centre [Dublin], *vol.1 nos.1-3 (1996)*.

Terminology *vol.3 no.1 (1996)*

The Translator *vol.2 no.2 (1996)*: Special issue on Wordplay & Translation.

Tribune des Industries de la Langue et de l'Information Electronique *no.20-22 (nov.1995-juil.1996)*. Special issue on Linguistic engineering and electronic information systems. Contents include sections on: Union Européenne, Francophonie, R&D, Produits, Utilisations, Stratégies, Entreprises.

Books

Akira Kurematsu and Tsuyoshi Morimoto: **Automatic speech translation: fundamental technology for cross-language communications**. Gordon and Breach Publ., 1996. (Japanese Technology Reviews, Section B: Computers and Comunications, vol.28) xiii, 111pp. ISBN: 2-919875-02-7. £34.00.

Anne-Marie Loffler-Laurian: **La traduction automatique.** Villeneuve d'Ascq (Nord): Presses Universitaires du Septentrion, 1996. 156pp. ISBN: 2-85939-502-4. FF.110.

The Translator's Handbook. 3rd edition, edited by Rachel Owens. London: Aslib, 1996. ISBN: 0-85142-352-3 (hb.) 0-85142-363-9 (pb.) Contents include: Terminology management systems (Klaus-Dirk Schmitz). -- Machine translation (Harold Somers and Clare Rutzler). -- New technology for tanslators (Geoffrey Samuelson-Brown)

Reports

European Commission: **Language and technology: from the tower of Babel to the global village**. Luxembourg, 1996. vi,25pp. (ISBN: 92-827-6974-7)

Conference proceedings

Expanding MT horizons. Proceedings of the Second Conference of the Association for Machine

Translation in the Americas, 2-5 October 1996, Montreal, Quebec, Canada. 289pp.

Items for inclusion in the 'Publications Received' section should be sent to the Editor-in-Chief at the address given on the front page. Attention is drawn to the resolution of the IAMT General Assembly, which asks all members to send copies of all their publications within one year of publication.