MT News International

Newsletter of the International Association for Machine Translation
Published in February, June and October

From the Editor

With this issue, *MT News International* enters its fifth year of publication. It has established itself as one of the principal sources of information about current developments in the field of machine translation and computer-based translation tools. With the aim of improving our service to IAMT members we are seeking your opinions and suggestions about the coverage and content of MTNI. Are there features and items which MTNI does not cover at present? Are there areas which it should not cover? Should it be confined strictly to news directly concerning machine translation and translation tools? Or should it continue also to cover other relevant developments in language resources, corpora, terminology, computational linguistics, lexicography, and natural

language processing?

In future issues it is proposed that MTNI will include regular columns from invited contributors, devoted to new products and systems, the activities of MT users and suppliers, and news from each of the three regional associations of IAMT.

In the next issue of MTNI (no.14 June 1996) we intend to include a questionnaire to solicit your views and suggestions, but we are of course grateful for feedback from our readers at any time (names and addresses of editors are on page 2).

NEWS FLASH
Michael Tacelosky Leaves Globalink

Effective December 1, 1995, Michael Tacelosky resigned his positions as a member of the Globalink Board of Directors and Technical Advisor to the President. Tacelosky, the founder of MicroTac Software, built the small San Diego software company into the leading producer of low-cost MT software. With MicroTac's success, MT software was available for the first time to retail customers at a price of less than $100.00.

Tacelosky became President of Globalink following the merger of Globalink and MicroTac, until Jim Lewis assumed the position in September 1995. He has always had an active interest in tobacco-control advocacy, and has decided to dedicate himself to these pursuits full time. Timothy Meekhof, Chief Scientist at Globalink, has assumed all technology duties previously performed by Tacelosky.

According to Tacelosky, "I moved to the heart of DC (Dupont Circle), and have taken a position with a the Advocacy Institute, a non-profit organization dedicated to sharing information with grassroots organizations who are doing advocacy work. I'm working with the tobacco control team, creating an Internet site to connect the hundreds of activists from around the country who work on a variety of tobacco projects: teen prevention, cessasion, regulation, etc.

Of course, I still love MT, and I expect to have integrate MT into the site within the next few months. (Latin America is a booming market for Philip Morris and their ilk, I'd like to assist the advocates there as much as possible)."

"The translation business has become a fast-growing, viable industry with mass market appeal, a development in which Tacelosky played a leading role. We wish him equal success as he turns his attention to philanthropic activities," said Harry Hagerty, Chairman of the Board of Directors. Muriel Vasconcellos, President of IAMT, added, "Michael Tacelosky has done more than any other single individual to popularize commercial machine translation. He introduced MT to the masses and reshaped the market so that there is no turning back: there is no longer any room for those who would hide it under a bushel and only trot it out on special occasions."

"Although I'm not working with MT development anymore, I will still be involved as an MT user, and I expect to see MT grow in the same way -- customized MT products on specific web sites, a natural domain-specific environment," added Tacelosky. Anyone wishing to contact Michael Tacelosky can reach him at his new e-mail address, MTAC@MSN.COM.

CONFERENCE REPORTS

Translating and the Computer 17
London, 9-10 November 1995

John Hutchins

The seventeenth annual conference in the well-known series organised by Aslib was held in the

imposing headquarters of the Institution of Civil Engineers close to the Houses of Parliament and Westminster Abbey. As in previous years, the conference attracted well over 200 participants from all parts of Europe and from a wide range of backgrounds in commerce, government and the translation profession.

The first day's proceedings, chaired by Veronica Lawson, opened with Gary Jaekel giving an account of the investment in MT made by the Ericsson Language Services, which delivers products and services in language training, translation and terminology to 23 companies in the Ericsson Group. The service deals with 15-20 thousand pages of telecommunications documentation per annum. The Logos MT system has been installed for about three years and the service reports a productivity increase of up to 40% with translations from English into French, Spanish and German.

Ursula Bernhard (GMD Forschungszentrum Informationstechnik) recounted experience with German-English and English-German translation using first the Logos system (1986-89), and then the METAL system (1989 until recently) to produce full translations. But now, since GMD researchers are encouraged to write in English, the need is for rough translations as drafts for producing documents, and the Translation Service at GMD is experimenting with tests involving Logos, METAL and the Globalink Power Translator. Ursula Bernhard reported initial results, finding that for German to English LOGOS and METAL were equally bad (or good) and that for English to German LOGOS was better than Globalink - although with longer use LOGOS and METAL have been more tailored to the specific needs of GMD than Globalink. Despite some poor quality, the majority of researchers considered raw MT could be a valuable drafting tool.

Dirk Lueke (SAP AG) described the experience at SAP with the METAL system since 1991 for the translation of manuals for its own business software. (SAP had earlier tried Logos in the mid 1980s but without success, primarily because neither SAP nor Logos were then fully ready for productive MT.) Lueke was able to report great success, with 7 translators currently producing on average 500,000 words per month. SAP is now looking for further applications of MT and more language pairs, to the integration of a translation memory in the setup, and to the provision of a single interface for all facilities. He stressed the importance of MT suppliers providing highly qualified support staff to ensure successful implementation, and the importance of MT users having greater influence on future MT product development.

Terence Lewis (Hook & Hatton Ltd.), speaking from experience with his own in-house MT system for Dutch-English translation and with tests of Globalink via Internet, argued that low-cost MT can achieve quality adequate for very many purposes, with the exception of documents intended for general publication. He claimed that with MT prices at 50% of those charged by freelance translators, industrial customers tend to lower their 'acceptability threshold', as long as the basic meaning of the original is conveyed.

In the following presentation, Svetlana Sokolova (PROject MT, St Petersburg) described the Stylus software for translation into and from Russian for English, French and German. Stylus is designed for PC running under DOS and Windows, it is currently the most widespread MT system in use in Russia, and has been purchased by a number of Western European and North American companies.

Also on the first day, Roger Harris spoke on the advantages and benefits to translators of using the Internet and electronic mail; Peter Kwan (Praetorius) looked at the significance for translators of UNICODE, an international code for scripts; and Joyce Pashley (SCO, UK) reported on the localisation of OpenServer 5.0 using the IBM TranslationManager.

The second day, chaired by Sigrid Martin, began with Carol Eckman describing the value of sharing and standardising terminology in the context of Norwegian government translation services. John Hatley of Logos then presented those contemplating the introduction of MT with a

set of guidelines and questions on the factors and costs to be taken into account. He stressed that MT does not replace translators, but reduces throughput time; translation itself does not make money, MT must be sold to company directors as a means of improving exports.

Gerda Klimonow (GMS) gave an insight into the linguistic problems of MT research, in this case the development of a Russian-German system for METAL. Based on many years' experience in the former GDR, the system is now nearing completion for the market in late 1996. It will be followed by work for German-Russian translation. Another research project was the topic of the paper by David Carter (SRI), this time the SLT system under development for spoken language translation at SRI. The system is designed for translation in the domain of air traffic planning (ATIS); it has been demonstrated in versions which translate spoken English into spoken Swedish and French, based on a vocabulary of 1200 items, and there is further development of versions for French into Spanish and from French and Swedish into English--thus allowing two-way dialogues.

Jose Verga (GSI-Erli) described the development and use of AlethTR at GSI-Erli as a system for large scale translation of internal documentation. This is an MT system of basic transfer design incorporating statistical methods, sophisticated terminology control and a translation memory. Although an internal tool, AlethTR will be available as a market product.

Robert Clark (Praetorius) spoke on the training of translators for the new technologies which they will encounter. At present, most do not acquire the necessary skills until in employment; it is clearly desirable that they receive some training at earlier stages, but in what form? The questions raised prompted a lively discussion.

The conference concluded with Anna Cordon (British Telecom), who described research at BT on various services for networks including a system for keyword searching of bibliographic databases, the NetSum system for producing automatic summaries of texts (and not just in English), already on trial to the public, and LiniText, a system for translating formulaic business correspondence (still under development).

During the conference, delegates were able to view demonstrations, either in the exhibition hall or in parallel sessions, of the translation workstations from Eurolang (Optimizer), Trados and IBM (TranslationManager), and also some of the systems described in presentations. In all, this was a successful and enjoyable conference, maintaining the high standards of previous events in this annual series.

The proceedings 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/aslib/)

Second Language Engineering Convention
London, 16-18 October 1995

John Hutchins

The first Language Engineering Convention was held by the European Commission in Paris in 1994. The second, held at the Queen Elizabeth II Conference Centre in London, was organised by the UK Department of Trade and Industry in conjunction with the Commission and sponsored by the Institution of Electrical Engineers and the British Computer Society. Its objective was to present to the business and professional public the products resulting from research supported by the European Commission in the field of language engineering.

A wide range of speech and language technology products were demonstrated in the exhibition hall and were described in presentations in the conference room. Among those included were aids for the disabled, spoken language communication with computers, authoring and localisation, information extraction and retrieval, speech synthesis, use and reusability of

language resources, and evaluation tools for language products.

Translation technology as such was specifically covered by accounts of the PaTrans system for translating Danish patents (Bente Maegaard and Viggo Hansen), of the SLT spoken language project at SRI (Manny Rayner), of the evaluation of MT software (Adriane Rinsche), and of the experience of SAP with MT (Jennifer Brundage). However, many of the other presentations had multilingual aspects of relevance to MT developments, particularly those concerned with language resource databases and with the establishment of evaluation methodologies.

It is an open question how much value this conference would have been for those not already familiar with language technology products and services. Many of the presentations were directed at researchers in the field or at the funders and sponsors, some outlining research in progress rather than existing prototypes. However, some of the plenary papers were succinct and masterly overviews of what had been achieved and what challenges lie ahead. Notable in this regard were the presentations by Karen Sparck Jones on the retrieval, extraction and summarisation of information and by Henry Thompson on why the achievements of language technology have been so frustratingly limited. He drew attention particularly to the fact that although high quality can be achieved when domains are narrow and restricted the resulting products have a small potential market. On the other hand, products with broad applications and hence potentially larger markets show poor quality performance and are likely to have low sales. It is essential that developers realise that language technology can only have a supporting role; they must abandon the goal of full automation and go for computer-assisted tools; they are in the business of cutting users costs at the margins, not large-scale redeployment of resources. His comments are highly relevant for MT, although many MT developers have learnt the lessons already by experience.

The third Language Engineering Convention will take place in Madrid in October 1996.

Software Localisation in Ireland – SLIG '95

Reinhard Schaeler

Ireland, one of the world centres for the localisation of software, now has its own Software Localisation Interest Group (SLIG) which held its first day-long Conference and General Meeting on 11th October at the University Industry Centre of University College Dublin. The conference was attended by representatives of most Irish based localisation companies, among them Microsoft, Lotus, Claris, Apple, Symantec, Intel, Berlitz and ITP, as well as researchers and members of the translation profession.

Ireland has established itself internationally as one of the major software localisation centres in the world, and the leading European location for such activity. Currently, 40-50% of the PC based software sold in Europe originates in Ireland. It is expected that this figure will rise to over 60% over the coming years. In Ireland, the localisation industry accounts for around 2 billion pounds worth in exports. There are approximately 4,000 people employed directly in the manufacturing and engineering of localised products. It is estimated that for every person directly employed two others are employed in dependent industries.

As part of the drive to maintain the momentum which has been built up in this sector, the National Software Directorate (NSD) proposed and facilitated the establishment of a special interest group for all parties involved in software localisation. This interest group, the Software Localisation Interest Group (SLIG), first convened in late February 1994, with representatives from academia and both overseas and indigenous companies attending.

On 11th October, SLIG held its first Annual Meeting and Conference at the University Industry Centre at University College Dublin. SLIG '95 was officially opened by Minister Pat

Rabbitte, T.D., who underlined the government's commitment to maintain and strengthen Ireland's role as the world's prime location for the localisation of software. Barry Murphy, the National Software Director, and Helen Wybrants, SLIG's chairperson, both highlighted in their opening remarks the remarkable work done by SLIG in the first year of its existence. They cited the active involvement in the preparation of proposals within the European Community's Fourth Framework Programme, activities within Ireland's own Strategic Research Programme, and SLIG's support for the setting-up of a Localisation Resources Centre under the Technology Centres Programme. Localising Multimedia Products, Education and Training for the Localisation Industry, Localisation Tools and Terminology Issues and Localisation were the topics discussed during the 4 main sessions of SLIG '95.

Seamus Gallen (FORBAIRT) chaired the first session which was introduced by Helen Wybrants (ISC-Europe) giving an overview of the SLIG's activities in the area of Multimedia. The first speaker, Paul Fenwick (Intel Ireland), focused on localisation of multimedia training materials. Marie Redmond (X-Communications), outlined the activities of her company which localises multimedia products for commercial markets. Marion MacDonald (Cormorant) concentrated on localisation issues which have to be addressed during the development of multimedia software.

The second session, chaired by Miriam Broderick (Dublin Institute of Technology), dealt with the educational and training requirements of the localisation industry. Introduced by Tony Mulqueen (Isocor) speakers came from a commercially run training institute (Alfie Keary, Centre for Advanced Technology Training – CATT), the Irish Translators' Association (Gabriele Milch-Skinner, ITA's Chairperson) and a 3rd level college (Richard Sutcliffe, University of Limerick). All three speakers presented the training programmes offered by their organizations and invited the localisation industry to become more actively involved in course development for specialized localisation training schemes.

Localisation Tools was the title of the third session, chaired by Paul Holden (Redacteurs Software Documentation). Reinhard Schaeler gave an overview of SLIG's Tools Group and announced that the negotiations with FORBAIRT to establish a Localisation Resources Centre at UCD – an initiative which has the backing of SLIG and a great number of individual localisation companies - was at a very advanced stage which envisaged an opening date of January 1996.

Further contributions to this session came from Orlagh Neary (Corel Corporation) who talked about her company's experience with machine translation, Orla Connolly (SSE) covering computer-aided translation for software localisation, John Rowley (Symantec) discussing the option of in-house vs. third party development of localisation tools and Alan Barrett (Lotus Development Ireland) outlining Lotus' involvement with the EU's 4th Framework Programme and its in-house development of a new tools suite for translators.

The session on Terminology Issues and Localisation was chaired by Deasun O Conchuir (NPi Training). Following the introduction of the SLIG Terminology Group by Marion Gunn (Everson Gunn Teo.), Jennifer Pearson (Dublin City University) reported on the European Terminology Initiative co-ordinated in Ireland by DCU, Anna Mazzoldi (Irish Translators' Association) gave an introduction to terminology resources on the Internet and Catherine Gavin (Berlitz) raised a number of important issues in relation to the industrial use of termbanks.

During the last session of the day, Helen Wybrants reported on the results of a survey carried out during the day among the participants of the conference. The objective of the survey was to give participants an opportunity to express their opinion on SLIG's first year of operation and to formulate strategies for future development. Participants had also been asked to indicate their interest in becoming more actively involved in one of SLIG's special interest groups or in SLIG's committee. The overall result of the survey showed a strong backing by those present for the ground-work performed by SLIG's outgoing committee. It also showed a very encouraging

interest among participants in becoming more directly involved through participation in one of SLIG's committees.

Helen Wybrants closed the conference thanking FORBAIRT, and especially the National Software Directorate, for their support in establishing SLIG and expressed her hope to be able to welcome the participants again next year, for SLIG '96.

Natural Language Processing Pacific Rim Symposium
Seoul, 4-6 December 1995

Ruslan Mitkov

The Natural Language Processing Pacific Rim Symposium (NLPRS'95) took place from 4 to 6 December 1995 in Seoul, Korea and was followed by a post-conference workshop on Language Engineering. The conference and workshop attracted researchers in various areas of Natural Language Processing and from many different countries (with the largest participation from Korea and Japan) and they were among the best organised events I have ever attended.

The number of papers accepted at the conference was 123 (68 regular papers and 55 poster papers) and their distribution in terms of countries was: Australia 1, Canada 1, Chile 1, China 9, France 3, Germany 2, Hongkong 2, India 2, Ireland 1, Japan 51, Korea 37, Malaysia 1, Singapore 1, Sweden 1, Taiwan 3, Thailand 1, UK 4, USA 3.

Of the above mentioned papers, 15 covered research and practical aspects of Machine Translation (6 papers from Japan, 5 from Korea, 3 from China and 1 from the UK). Machine-aided translation architectures were described in the papers "DEAR: A translator's workstation" (Ming Zhou et al) and "Interactive machine-aided translation reconsidered – interactive disambiguation in TWP" (Kiyoshi Yamabana et al.). Pattern-based MT was a hot topic at NLPRS'95 and was presented in the papers "Pattern-based MT: accomplishment of the BT863 system" (Zhao Tiejun et al.), "Patterns from large corpora for Korean-English MT" (Nari Kim e al.) and "A method for target word selection using similarity matching for MT" (Kwon Yang Kim et al.). Discourse issues for MT were discussed in the papers "A discourse analysis model using hybrid knowledge for dialogue MT" (Jae Won Lee et al.) and "Text-wide MT grammar"(Jiri Jelinek et al.), whereas more traditional rule-based approaches were reported in "A hierarchical intelligent rule-based system (Xue-qi Cheng et al.) and "Changing syntactic classes in transfer-based MT (Yoshihiro Matsuo et al.). In addition, the memory-based translation program STAR was demonstrated at the conference.

Being predominantly an Asian conference (but this time with large world-wide participation), NLPRS'95 confirmed the Asian interest in MT, leaving the impression that researchers in Asia tend to experiment more and more with non-traditional (non rule-based) MT methods.

NLPRS'95 statistics were kindly supplied to me by the Program Chairman Prof. Hozumi Tanaka and by the Organising Chairman Prof. Key-sun Choi.

SYSTEMS and PRODUCTS

Language Engineering Corporation announces new version of LogoVista E to J and two new services

[From press releases, 3 November 1995]

Version 2.5 of LogoVista E to J. The new release of LogoVista E to J system is available for Macintosh, Power Macintosh, Microsoft Windows 95, Windows 3.1 and Windows NT 3.5 or later. It features also a revised and expanded dictionary (now 120,000 entries), twenty-one optional technical dictionaries (together containing more than 540,000 specialized terms for business, science, and technology), and new translation memory archives. The new release is being sold at $995 for both Macintosh and Windows versions (upgrades for $350).

Translation of WWW pages. Language Engineering Corporation has announced a new service for translating English World Wide Web pages into Japanese. A special version of LogoVista E to J preserves the HTML markup during the translation process, keeping all images, hypertext links and other HTML features in place. Clients may choose either to have the translated pages delivered without postediting or to have the postediting done by Language Engineering Corporation, who will deliver finished pages ready to install on Web servers. Prices for postedited pages range from $300 (for 200 words) upwards; for HTML translation without postediting the charge is $0.10 per word (minimum charge $100).

Electronic mail from English into Japanese. LEC has also announced a new service for translating English into Japanese via e-mail. A special version of LogoVista E to J receives English e-mail messages, translates them into Japanese, and returns the translated messages to the sender by e-mail. Customers may specify the technical dictionaries to be used. Response time depends on the speed of the e-mail and can be as little as a few minutes. The e-mail service is priced at $0.03 per English word, with a minimum charge of $5.00.

For further details on all these developments, visit LEC's World Wide Web site (http://www.lec.com), send e-mail to info@hq.lec.com, or call (800) 458-7267 or (617) 489-4000.

Systran Announces Web Page Translator

[Press release]

One of the most exciting phenomena of the 1990's has been the World Wide Web (WWW), the hyperlinking of computers and users throughout the world. The Web is truly international: Web users can *surf* a site in Berlin one minute, New York City the next, and Turkey moments later. The language barrier persists, however, with translations of Web pages being provided only on certain rare occasions, and only for selected pages.

Systran Software, Inc., a leading developer of Machine Translation (MT) software, now offers a free service for multi-lingual surfers: a Web Page that allows users of the Web to translate any page on the Web to or from English. Soon to support all of Systran's many language pairs including Japanese and Chinese, the service currently provides support for Russian-English, bi-directional Spanish, French, German, Italian, and Portuguese. At http://systranmt.com/ translate.html, all a user need do is provide the address of the page to be translated, the language pair involved, and the user's E-mail address. After a short wait, the desired Web page in its machine-translated form will be displayed on the user's browser.

Useful both for those from non-English speaking countries wanting to surf the largely English Web, or for those adventurous English speakers wishing to explore exotic distant sites, this service provides users with a taste of the new, electronically linked world.

Globalink Launches On-Line Professional Translator Service

[Press release]

Globalink, Inc. has unveiled in the next phase of its corporate strategy to provide customers with a full range of translation solutions. The new service, Translate Direct, links customers

electronically to Globalink's network of more than 1,000 professional translators worldwide, available via Internet or modem.

"Most people would agree that locating a freelance translator on short notice, who can turn around a project promptly with good quality, is virtually impossible. In fact, the translation process has been so tedious and costly that many businesses avoid it, suffering the consequences of missed opportunity simply because they don't speak a foreign language. Translate Direct is able to provide a fast, cost-effective and convenient solution, directly from any PC, taking the hassle out of translation."

To access Translate Direct, Globalink provides free interface software; a client version is sent to customers, while carefully screened professional translators receive their own version. Customers identify a file to be translated and choose the source and target languages from a menu. Based on the level of service selected, the program quickly calculates a price and time estimate--ranging from fast and inexpensive machine translation using the Barcelona technology, to more expensive publication-quality work by a professional translator. Customers can easily establish a corporate account or pay by credit card and then send the document to Globalink by modem or Internet e-mail, as well as fax and mail. Customers are notified by fax, phone, or e-mail upon job completion.

"Globalink gives customers a full range of translation options--allowing them to determine which method of translation best suits their needs and budget. Globalink selects the most appropriate translator for the job, and ensures that it is completed on time and meets quality standards before returning it to the client, either electronically, by fax, or even by mail," commented Jim Lewis, Globalink's president. For Professional translators, Translate Direct offers a steady work load and eliminates time-consuming administrative and accounting procedures.

For more information on Globalink's services or products, and to receive a free Translate Direct disk, call 800-255-5660 or visit our Web site at http://www.globalink.com.

PARS/U For Windows
World's First Commercial English-Ukrainian-English
Machine Translation System

Michael S.Blekhman

1. Background

Ukraine is a large country situated in the middle of Europe, between Poland and Hungary, on the one side, and Russia, on the other side. We gained our independence on December 1, 1991, and since then, the language problem is Problem Number One here, being even more vital for many people than the low standards of life. The fact is that Russian has long been the only state language in ex-Union, and the number of native speakers of Ukrainian is not large; the reasons are numerous, one of them being the mass-scale extermination of language-bearers in 1930s.

It's only natural that the serious politicians in the Parliament insisted on the Law on the languages, adopted several years ago, according to which Ukrainian is the only state language in Ukraine. It turned out immediately that mastering another language was a serious challenge for the majority. Though Ukrainian and Russian are rather close, they are different languages that reflect different cultures.

And, while English-Russian lexicology has a great tradition here, English-Ukrainian lexicology has almost no tradition at all. This means that the first task to be solved is compiling English-Ukrainian and Ukrainian-English dictionaries.

Another task, whose importance is by no means smaller, is creating Ukrainian-oriented language engineering, which means, among other tasks, developing machine translation systems

to and from Ukrainian.

Language engineering traditions are very rich in ex-Union, so it was not hard to develop the first MT products. It's no surprise that the first such system, RUMP, translated texts between Russian and Ukrainian. It was developed in 1991 by Lingvistica '93 Co (named Medicom Ltd at that time, and later, PARS Ltd).

At present, Ukraine is being drawn into the world scientific and business communities, and the need for commercial MT systems translating between Ukrainian and main European languages, English first of all, is very strong. Dozens of people and institutions have been asking me when a system like PARS (the well-known English-Russian-English MT package) will be developed.

In January, we started marketing PARS/U, which, we hope, will be as broadly used in Ukraine as PARS.

2. Problems

The first problem to be solved was lexicographical. We have developed a bidirectional English-Ukrainian basic dictionary of 33,000 words and phrases, which contains not only general-usage, but also some business terms. When compiling the dictionary, we have made use of some English and English-Ukrainian printed dictionaries, and we have also run numerous original texts through PARS/U to enlarge the dictionary.

Our next step in this direction will be creating a very large general dictionary based on Webster's lexicon, which is now being done for PARS. We are also developing terminological dictionaries for PARS/U. A dictionary of computer terms is under way, which will be based on several dictionaries on information science compiled by Ukrainian lexicographers. The polytechnical and ecological dictionaries are also being developed.

Another problem was working out a formal grammar of Ukrainian. We have made a thorough description of Ukrainian morphology and developed a Ukrainian grammatical dictionary. Dozens of rules for analysis and synthesis of Ukrainian texts have also been written and included into PARS/U.

Certainly, PARS is the system behind PARS/U, but numerous Ukrainian peculiarities were taken into consideration, such as 7 morphological cases in Ukrainian (Russian has 6), differences between Ukrainian and Russian participial forms, etc.

3. How it works

PARS/U runs on IBM PCs under Windows. It is compatible with any Windows-supported text processor, such as Word Perfect or AmiPro, via Clipboard. The user may mark a text portion, copy it to Clipboard, activate PARS/U, have the text translated, and copy the translation back from Clipboard to his/her text processor.

Besides, PARS/U is embedded into MS Word 6.0 for Windows. When the latter is activated, the "Translate" item is displayed in its main menu, which lets the user activate any of Lingvistica's Windows-based MT systems: PARS/U, PARS for Windows, or RUMP for Windows. The user can open an English or Ukrainian text in WinWord and have it translated. The target text will appear under the source one, in a separate window activated automatically by WinWord. The source text format is fully preserved in the target text, including fonts, styles, paragraphs, tables, etc.

PARS/U features several options, which, I suppose, are to be available in any commercial MT product.

a) The polysemantic words are marked with asterisks in the target text, so that the user could select a more appropriate translation if necessary. A double click on an asterisk displays a panel of variants; another double click will substitute the initial word in the text with the word (phrase) selected by the user.

b) The user may enter a word or phrase into the dictionary directly from the screen by

marking it in the WinWord window and clicking on the "New word" item in the PARS/U menu; the word (phrase) will be immediately inserted into the dictionary. This has proved a reliable and convenient way of customizing PARS/U. The user runs a representative corpus of his/her texts through PARS/U and creates a dictionary of specific terminology.

c) PARS/U features the "Auto-encoding" routine. It assigns grammatical characteristics to the Ukrainian words entered into the dictionary. When analyzing a word, PARS/U uses the Ukrainian grammatical dictionary. Taking into account the high morphological ambiguity of Ukrainian words, PARS/U displays variants in those situations when it cannot choose between two variants, for example, when the noun/adjective ambiguity is encountered. In such situations, it is up to the user to make the final choice.

d) PARS/U also features the so-called batch operations with dictionaries. In particular, the user may have 2 dictionaries merged, which lets him/her share dictionary compilation between several dictionary officers, and make one dictionary out of the several ones.

e) PARS/U dictionaries are bidirectional: creating, for example, a Ukrainian-English dictionary, the user obtains a Ukrainian-English-Ukrainian one.

f) The last but not the least fact to mention is that PARS/U runs both in stand-alone and network modes.

4. How it is delivered

PARS/U is supplied both on diskettes and on CD-ROM. In the latter case, the compact disk contains the following products:

a) PARS/U;

b) PARS for Windows;

c) PARS for DOS;

d) PARS Tutor: a Windows-based tutorial program in two versions, in English and in Russian, with 100 illustrations in each version; the tutor lets the user master working with the above-mentioned MT systems;

e) PARS/DOS Demo version for the users running PARS under DOS;

f) RUMP for DOS.

For more details: Dr. Michael Blekhman, 94a Prospekt Gagarina, apt.111, Kharkov 310140, Ukraine. (Tel: (0572) 27-71-35; Fax: (0572) 40-06-01; Email: blekhman@lotus.kpi.kharkov.ua.)

CORPORA and SERVICES

Localisation Resources Centre in Ireland planned for 1996

Plans to open a Localisation Resources Centre (LRC) in Ireland are at an advanced stage. The opening of the Centre is now planned for January 1996. A proposal by the Department of Computer Science of University College Dublin (UCD) to FORBAIRT, Ireland's Science and Industrial Development Agency, has been approved. Negotiations to sign an agreement under which the new Centre will operate will more than likely be finalized by the end of this year thus opening the way for the establishment of probably the first centre of its kind in Europe.

Irish based localisation companies have over the past 2 years carried out expensive market research and evaluation exercises to identify suitable 3rd party tools. It is widely accepted in the industry that in order to avoid unnecessary duplication of efforts and to make optimal use of existing resources, this work should be carried out by one independent centre. The LRC will address the urgent need to gather and disseminate all available (but sometimes not easily accessible) information on tools and to develop new tools according to the requirements of the localisation industry.

The Localisation Resources Centre will offer the following services to the industry:

- o Research and Development of Localisaton Tools
- o Evaluation of Localisation Tools
- o Localisation Tools Library
- o Education and Training
- o Consultancy
- o Industry Watch
- o Regular Publications

The setting-up of the Centre will be funded under FORBAIRT's Technology Centres Programme. After an initial start-up period, it is planned to transform the Centre into a limited company which will offer its services to publishers and localisation service providers.

For more information please contact: Reinhard Schaeler, Department of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland (email: reinhard@nova.ucd.ie; tel: +353-1-269 3244 ext.2472, or direct: +353-1-706 2486; Fax: +353-1-269 7262)

English/German Translation Corpus

We are presently compiling an English-German, German-English translation corpus. We have so far concentrated on the English-German part, of which over half a million words have been computerized to date. The English-German part is divided into a core corpus and a subcorpus. Core corpus parameters include British English, written and non-literary, with the texts ranging from academic textbooks from various domains (e.g. history, philosophy, the arts, economics and physics), to publications by the European Community/Union and a selection of tourist brochures. Our subcorpus consists of two categories which deviate from the main parameters of the core corpus, namely contemporary British literature on the one hand and scripted public speeches - language which is written to be spoken - on the other. Since it is not always possible to find parallel categories for the two directions, the German-English part will at least consist of academic textbooks, literature and tourist brochures. We would be grateful for ideas and suggestions on what else we could include in our corpus. Since copyright problems are a major obstacle to corpus compilation, we are particularly interested in material which is freely accessible and perhaps even available in machine-readable form. More specifically, are there any newspapers that publish translated articles regularly, or are there any special translation pages on the internet? More information on the project will be available on the internet shortly under http://www.tu-chemnitz.de/~ehe/public/real.htm

Prof. Dr. Josef Schmied, English Language and Linguistics, Reichenhainer Str.39/223, TU Chemnitz-Zwickau, 09107 Chemnitz (Tel.: +49 371 531-4226; +49 371 531-4279; Fax.: +49 371 531-4233; Email: Josef.Schmied@phil.tu-chemnitz.de)

Language Software Helpdesk at HCRC

The Human Communication Research Centre is pleased to announce the launch of the Language Software Helpdesk: a free service dedicated to the support of public domain and freely available software for natural language processing and the fostering of its use in practical applications.

There is a very substantial body of software available in the area of natural language processing, both in the form of complete systems and of system components. But the uptake of this software, particularly by industrial users, has been impeded by lack of information about functionality and worries about the availability of ongoing support. The Language Software Helpdesk is intended to directly address both these problems.

Contact: Language Software Helpdesk, c/o Language Technology Group, Human Communication Research Centre, 2 Buccleuch Place, Edinburgh EH8 9LW, UK (Fax (marked

'Language Software Helpdesk'): +44 131 650 4587; Email: Language.Software.Helpdesk @ed.ac.uk)

BNC Word-POS frequency lists available

Frequency lists for English word-POS pairs, produced from the British National Corpus, are available for ftp (username: "anonymous") from our ftp site, ftp.itri.bton.ac.uk in directory pub/bnc

Frequency lists are available for:

cg 'context-governed' spoken material: 6.2M tokens, 79,906 types
demog 'demographic' spoken material (eg conversation): 4.2M tokens, 54,652 types
written: 89.7M tokens, 921,074 types
all: 100.1M tokens, 939,028 types

They are all available in 2 forms, sorted alphabetically ("al") or by frequency (highest frequency first) ("num"). They are compressed using gzip. The frequencies are for CLAWS-word, CLAWS-POS pairs.

For further information on the BNC see http://info.ox.ac.uk/bnc

Lexicons from Merriam Webster and the Linguistic Data Consortium

The LDC has reached an agreement with Merriam Webster, Inc. to distribute electronic versions of Merriam Webster's Collegiate Dictionary, Tenth Edition (full text), Biographical Dictionary, New

Geographical Dictionary, Collegiate Thesaurus, School Dictionary, and The Merriam-Webster Dictionary (mass-market paperback) to the research community. Except for the Collegiate Thesaurus, each lexicon will be available in two versions, full text and a version with headwords, functional labels, and pronunciations.

The cost of the lexicons ranges from $100 to $300 for LDC nonmembers; LDC members will have the option of receiving one copy of each lexicon as part of their 1996 membership. A special agreement between Merriam Webster and the user will be signed prior to releasing the lexicon. 'Research Only' licenses are available from the LDC; requests for commercial licenses will be directed to Merriam Webster.

Questions and requests for obtaining any of the Merriam Webster Corpora should be sent to ldc@unagi.cis.upenn.edu.

Second Release of CELEX CD-Rom

The Second Release of the CELEX CD-ROM is announced by the Dutch Centre for Lexical Information and the Linguistic Data Consortium.

This CD-ROM contains an enhanced, expanded version of the German lexical database (2.5), featuring approximately 1000 new lemma entries, revised morphological parses, verb argument structures, inflectional paradigm codes, and a corpus type lexicon. A complete PostScript version of the German Linguistic Guide is also included, in both European A4-format and American Letter format. For German, the total number of lemmas included is now 51,728, while all their inflected forms number 365,530.

Moreover, phonetic syllable frequencies have been added for (British) English and Dutch. Apart from this, and the provision of frequency information alongside every lexical feature, no changes have been made to the Dutch and English lexicons.

Complete AWK-scripts are now provided to compute representations not found in the

(plain ASCII) lexical data files, corresponding to the features described in the CELEX User Guide, which is included on the CD as well.

For each language, i.e. English, German and Dutch, the CD-ROM contains detailed information on the orthography (variations in spelling, hyphenation), the phonology (phonetic transcriptions, variations in pronunciation, syllable structure, primary stress), the morphology (derivational and compositional structure, inflectional paradigms), the syntax (word class, word-class specific subcategorisations, argument structures), and word frequency (summed word and lemma counts, based on recent and representative text corpora) of both wordforms and lemmas. Unique identity numbers allow the linking of information from different files with the aid of an efficient, index-based C-program.

Like its predecessor, the CD-ROM is mastered using the ISO 9660 data format, with the Rock Ridge extensions, allowing it to be used in VMS, MS-DOS, Macintosh and UNIX environments. As the new release does not omit any data from the first edition, the current release will replace the old one.

Non-members of LDC can receive a copy of CELEX for research purposes only for a fee of $150. Information: ldc@unagi.cis.upenn.edu, or Fax: (215) 573-2175; Tel: (215) 898-0464.

Web Site for SALT

The UK Speech and Language Technology Club has now set up a website for obtaining information about SALT and its activities. URL address: http://salt.essex.ac.uk/salt/

The new website has pages referring to topics of interest and work being done outside the UK, especially in the European Union.

Those who regularly post notices on the SALT bulletin board, please note that the new address: salt@essex.ac.uk

SECC available on WWW

Gert Adriaens

Here is some news that will certainly interest people active in the field of controlled languages:

* The European SECC project (A Simplified English Grammar and Style Checker/Corrector) offers the possibility of trying out the SE tool it is developing via WWW. You can submit a sentence and get "network-instantaneous" feedback about errors, as well as a proposal for correction. The URL is http://www.ccl.kuleuven.ac.be/cgi-bin/seccdemo.cgi

* Willem-Olaf Huijsen of the University of Utrecht, the Netherlands, has started working on a "Controlled Languages Home Page", and welcomes any input. The URL is http://wwwots.let.ruu.nl/Controlled-languages/

Hybrid Models: Mailing List

As announced at the CSI workshop at IJCAI last August, a new mailing list has been set up for the specific purpose of exchanging information and ideas regarding hybrid models, especially models integrating symbolic and connectionist processes. Other hybrid models, such as fuzzy logic+neural networks and GA+NN, are also covered. This is an unmoderated list. Email submissions to hybrid-list@cs.ua.edu. To subscribe: send an e-mail message to the userid "listproc@cs.ua.edu" with NO SUBJECT, but the one-line text message: SUBSCRIBE hybrid-list YourFirstName YourLastName

MT Course at UCLA

On April 22-24, 1996, UCLA Extension will present the short course, "Machine Translation", on the UCLA campus in Los Angeles. The instructors are Eduard Hovy, PhD, USC Information Sciences Institute; Kevin Knight, PhD, USC Information Sciences Institute; and Jaime Carbonell, PhD, Carnegie Mellon University.

Machine translation describes computer translation of one human language to another, and is one of the oldest large-scale applications of computer science. The need for systems to translate documents to and from a variety of languages is expanding, for applications as diverse as:

- o Multilingual e-mail
- o Browsing (such as on the World Wide Web) texts in other languages
- o High-quality translation of business letters and reports
- o Translation of technical documents and articles
- o Speech-to-speech translation for business and travel.

While useful MT technology is currently available, it is not yet capable of providing both high-quality and wide-domain performance simultaneously. For higher quality, the domain may be limited, and human assistance required while for wider domain, output quality may be sacrificed. MT research continues to push the boundaries of this automation-quality-scope continuum. New techniques, such as statistical MT and example-based MT, add new capabilities and possibilities to the older tried-and-true methods and theories of MT. But comparing systems, and measuring MT quality, can be challenging.

This course covers the entire scope of machine translation, including the original and the latest techniques and technology. It is intended for both the interested layperson as well as the computer science professional who wants to become familiar enough with the technology to construct a simple MT system, or to make informed decisions when purchasing an MT system or MT services.

The course fee is $1195, which includes 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

PUBLICATIONS REVIEWED, ANNOUNCED and RECEIVED

Ovum's Translation Tools Report

Colin Brace

[Part one of a two-part review]

British consultancy and market research firm Ovum's *Translation Technology Products* is an ambitious survey of commercial translation software. The hefty A4 size report is handsomely produced and is abundantly sprinkled with tables, flowcharts, graphs, and listings. Ovum, which publishes numerous such reports each year, clearly knows its audience: corporate managers who need clear, concise information on a hitherto unfamiliar field in a hurry -- and are willing to pay for it.

Over the years, Ovum has published a variety of studies on the field of language technology. *Translation Technology Products*, however, is different in at least one respect: this time around Ovum takes a stab at evaluating products. The report's authors, language industry veterans Jane Mason and Adriane Rinsche, evaluated five MT systems (DP/Translator,

PowerTranslator Professional, Logos, LMT, Metal, and Systran), five translator workbenches (Eurolang Optimizer, XL8, Translator's Workench, Transit, and Translation Manager), and two terminology management programs (MTX and Translexis). In addition, they provide a directory of another sixteen companies offering software which the authors have not evaluated.

As was the case with its 1991 study, *Natural Language Markets*, Ovum does not include the Japanese market, despite the large number of translation systems developed and marketed there. Ovum offers no explanation for this; one can only conjecture that it may be due to practical reasons, such as lack of ready access to a Japanese linguist, or due to strategic considerations, such as the belief that such coverage will not improve sales of the report.

Translation Technology Products is broken down into the following sections:
* management summary
* What is translation technology?
* Using translation technology
* Market scenario and forecasts
* Introduction to the evaluations
* Evaluations: Machine translation tools
* Evaluations: Translator's workbenches
* Evaluations: Terminology products
* Directory

The first three sections offer a straightforward tutorial for readers with no previous knowledge of translation technology. More controversial is Section D, Market scenario and forecasts, which offers the obligatory (for Ovum) market forecasts. The authors write: "In 1995, we estimate the value of the translation products market to be US$ 200 million worldwide. Over half of that market is for automatic translation products (US$ 112 million), a large proportion of which are Japanese. These figures include all language pairs in all markets. The market will grow to US$ 1.5 billion by 2000." (These figures are based on a previous Ovum report: *Globalisation: Creating New Markets With Translation Technology* [see MTNI#12]) Ovum's projection of US$ 200 million for 1995 sounds high and one wonders how it arrived at this figure, for its own published data does not bear this optimism out. In their report, the authors list the following revenues in 1994 for the vendors surveyed:

Eurolang: FFR 4-6 million (US$ 0.75-1.0 million)
Globalink: US$ 17 million
Globalware US$ 480,000
IBM: US$ 1 million
Intergraph: US$ 1 million total revenue since 1989
LinguaTech: DM 300,000 (US$ 200,000)
Logos: US$ 1.5 million
Sietec: DM 9 million (US$6 million)
Star: DM 675,000 (US$ 450,000) total revenue since 1992
Systran: US$ 2.5 million
Trados: DM 1.3 million (US$ 870,000)

This gives us a total US$ 30 million. It is possible -- though highly unlikely -- that Japanese vendors may have achieved a similar turnover in 1994, in which case total worldwide volume might conceivably have reached US$ 60 million for 1994. While Ovum assures us the above vendors were expecting substantial growth for 1995, it is difficult to believe that this already impressive market in turn nearly doubled in size from 1994 to 1995.

This being said, estimates on size of the translation market are extremely hard to come by; Ovum is virtually the only organization in the world attempting to survey it. For those in urgent need of such data, Ovum is currently the only game in town.

The quality of the evaluation sections of the report will be analyzed in detail in the second half of this review, which will appear in the next issue of MTNI (no.14, June 1996). It will suffice to note here that each vendor is evaluated separately and the evaluation encompasses a variety of information. Of particular note are the discussions of commercial viability, what Ovum describes as company's "competitive position." Some of the most engaging reading in the report is found in these company profiles; this is an area where previous books on MT have systematically fallen short. For example, the Ovum authors, clearly impressed with the growing array of offerings of the world's biggest computer company, write:

"IBM has the motivation and desire to be a leading vendor in the translation technology market and is able to provide financial stability to the science centres and software development labs. However, the success of LMT on a worldwide basis will depend on the availability of additional language pairs, the development of a clear marketing strategy, and adequate distribution channels."

Mason and Rinsche can nonetheless be charmed by more modestly sized companies as well:

"Trados is a very small and dedicated company focusing on innovative integrated translation technology products. The directors are young, sharp, and dynamic.... revenue [levels] are impressive from a company of only five employees. This reflects the level of interest shown in the innovative techniques adopted by the Workbench and MultiTerm. Although small, Trados appears financially stable, with excellent prospects for the future if adequate marketing channels and support can be developed for the US."

Pondering the latest moves of the elder statesman among MT developers, the authors write:

"...the reason Systran is still around after so many years is its management's approach to commercial viability. Systran has traditionally focused on multinationals producing technical documentation, in areas such as automotive, engineering, science, and technology, where significant volumes and tight deadlines are a major concern. However, it is looking at new areas, including language learning and education, multilingual data retrieval, and other natural language processing projects."

and: "Systran is considering the mass marketing of Systran by introducing a standalone, PC-based version, clearly placing the product in a highly competitive market. Systran has yet to determine the marketing channels for such a product. However, Ovum believes Systran has the experience and the commercial focus to make an impact in this sector of the translation technology market, should it decide to do so."

Since Systran has been converted to Windows and is available on cheaper and more open hardware platforms, Ovum believes that Systran is right to be optimistic about its prospects.

Market prognoses aside, Ovum's *Translation Technology Products* is a well researched and pleasantly readable compendium of product and company information. Regrettably, the exceedingly high price of the report is likely to prevent it from enjoying the wider readership it deserves within the MT community.

Publication details: **Ovum evaluates Translation Technology Products** [by] Jane Mason [and] Adriane Rinsche. June 1995. 424 pages. Price: #995/US$1850. ISBN: 1-898972-20-6. Publisher: Ovum Ltd., 1 Mortimer Street, London W1N 7RH, UK (Tel: +44 171 255 2670; Fax: +44 171 255 1995)

International Who's Who in Translation & Terminology

John Hutchins

A group of four organisations has just published the first edition of what may well become a

standard reference work. The *International Who's Who in Translation & Terminology* is the first publication to give biographical details of leading figures in the language professions of translation (including machine translation), interpretation and terminology. The idea of compiling this who's who originated from talks between Geoffrey Kingscott of Praetorius Ltd. and Karl Baddé of Where+How International; they were soon joined by Union Latine, which had had similar plans, and later by Infoterm, which already had a database of terminologists. The four organisations agreed to work together, consulting and gathering data from as wide a range of contacts as possible. Given the great variety and international spread of the domains involved, it is inevitable that, as a first attempt, this cannot be comprehensive. To a large extent, compilers of such publications rely on the cooperation of their respondents. It is measure of the efficiency and diligence of the compilers that so many whom one would hoped to find in such a who's who are in fact present.

The details provided for entrants include nationalities, dates of birth, addresses and telephone/fax numbers, qualifications, fields of activity and specialisation, languages, current positions, memberships of professional bodies, and selected publications. The volume is prefaced by Jean-François Joly of the Féderation internationale des traducteurs, and introduction and descriptions of the four collaborating organisations - each in four languages: English, French, Spanish, and German. The publication comes also with a diskette (for IBM PC compatibles) containing indexes (in four languages) for the subject expertise and language knowledge of all those included.

In sum, this is a most useful who's who, which fills a major information gap and which, it is very much to be hoped, will become a standard reference work for all involved in translation and terminology as it grows in coverage and value with every new edition. The compilers are to be congratulated for their initiative and for the success of this first edition.

Copies are available from: Praetorius Ltd., 128 Derby Road, Long Eaton, Nottingham NG10 4ER, U.K. (£72 per copy); Where+How North America, PO Box 16162 Beverly Hills, California 90209-2162, USA ($115 per copy); Rod Churchill-Bateman, International Where+How, PO Box 36, Lyneham, ACT 2602, Australia (AUD$160 per copy); Union Latine, 131 rue de Bac, 75007 Paris, France (FF660 per copy); and International Where+How, Postfach 2464, 53113 Bonn, Germany (DM180 per copy).

The Text Encoding Initiative: Background and context. Edited by Nancy Ide and Jean Veronis. (Kluwer Academic Publishers, Dordrecht). 230pp. Hardbound (ISBN: 0-7923-3689-5) Hfl.145, $99, £60. Paperback (ISBN: 0-7923-3704-2) Hfl.75, $45, £29.

The Text Encoding Initiative (TEI) Guidelines for Electronic Text Encoding and Interchange are the result of over six years' work by dozens of scholars from all over the world. As such, they represent a pioneer effort in an area where only occasional and isolated attempts were made before. They will certainly serve as the primary basis for encoding texts in electronic form for the foreseeable future.

The work of participants in the TEI not only involved consideration of problems of text encoding that are likely to be with us for decades to come, but also required the development of a methodology - from scratch - for approaching these problems. These pioneering efforts, while likely to be refined and extended, must not be lost: they provide the intellectual basis upon which text encoding practices will build in the future. This collection documents the course of these efforts.

"The TEI Guidelines are extraordinary. Even if they were never adopted they would stand as a significant contribution to scholarship for their detailed analysis of the information sets of a huge range of complex text types." From the Preface by Charles E. Goldfarb, inventor of the

Standard Generalized Markup Language.

Publications from Localisation Industry Standards Association

LISA announces the availability of the LISA Quality Assurance Model and the LISA Showcase.

The *LISA Quality Assurance Model* is designed to assist the localisation research, production and quality control processes for software developers and service vendors. The LISA QA Model provides guidelines and metrics for localisation quality. The documentation and templates help clients and vendors implement a synchronised QA system at multiple production sites. All registered users will be published by LISA to familiarize more localisers with the model and to increase networking. LISA Membership Price $ 175; Non-Member Price $ 275

The *LISA Showcase* is an extensive reference tool containing detailed information about the products, processes, standards and methodologies for the localisation and internationalisation business. It is the most comprehensive guide to localisation and language processing tools available today. The LISA Showcase is a decision making tool for anyone involved in today's localisation business. It is now available on CD-ROM and is a directory and catalogue of localisation and language processing tools, supplier profiles, products, and guidelines. It also contains industry trade journals, independent product reviews and European Community localisation and software development project reports. Published twice yearly, the *LISA Showcase* will provide precise and up-to-date information. It will run under Windows on standard hardware with a state-of art user interface that will cross- reference data for easy information access. LISA Membership Pricing: 1 CD $150.00; 3 CDs $395.00; 5 CDs $625.00; 10 CDs $975.00; Non-Member Price each CD $250.00.

If you would like further information about ordering either of the products contact: Jane Barney, LISA Administration, 2bis rue Ad-Fontanel, CH-1227 Carouge, Geneva, Switzerland (Tel: +41 22 301 5760; Fax: +41 22 301 5761; Email: manobile@divsun.unige.ch

Computers and the Humanities
Special Issue on Computer Programming for the Humanities
Guest Editor:
Eric Johnson <JohnsonE@columbia.dsu.edu>.

Topics: Submissions of articles are invited that focus on any aspect of computer programming for the humanities -- including articles on topics such as the following:

* Programming methodologies and software design principles used to create computer programs in the humanities; in particular, description of facets of humanities programming which distinguish it from other kinds of programming;

* An overview (or projection of the future) of programming for the humanities using C and C++, Icon, Pascal, Perl, SNOBOL4 and SPITBOL, other computer languages

* Descriptions of actual programming experiences (recently finished or in progress) which raise significant questions and problems.

* Description of a specific programming application (or a type of application) for the humanities -- including the visual arts, drama, history, and music as well as literature and linguistics.

In addition to technical papers, general discussion or opinion papers are invited on topics that grapple with questions such as the following:

* Do humanists who create computer programs do so in ways different from computer scientists? Do they more (or less) readily grasp an overview of a computing problem and see the general framework of a solution? Do they write computer code differently? Do they prefer

particular computer languages?

* Occasionally those with humanities educations and solid academic positions in the humanities assume positions normally held only by those with degrees in computer science. How is that possible? Do those with educations in computer science ever assume positions in the humanities?

Length: Articles of any length will be considered. It is expected that articles will range from 2,500 to 12,000 words -- except for opinion articles or overview articles which might be shorter.

Format: Submissions should be in the same form as regular submissions to Computers and the Humanities:

Submission: All submissions should be via electronic media -- email and FTP are strongly encouraged. Articles that can be saved as ASCII files (with line breaks and lines no longer than 80 characters) should be sent via email to the guest editor, Eric Johnson, at JohnsonE@columbia.dsu.edu

The guest editor should be contacted via email at the above address about arrangements to transmit articles containing special characters or graphics that cannot be saved as ASCII files.

Deadline: March 15, 1996

Computational Linguistics
Special issue on Word Sense Disambiguation.
Guest Editors:
Nancy Ide <ide@cs.vassar.edu>, Jean Veronis <veronis@univ-aix.fr>

The discrimination of word senses, word sense disambiguation (WSD), is of prime importance for all areas involving computerized language analysis, including corpus-based research, lexical studies, information retrieval, machine translation, natural language processing, studies of style and theme, authorship attribution, and applications such as hypertext browsing.

This special issue is intended to summarize the state of the art in WSD, identify the main areas of difficulty, and suggest solutions for improvement. Papers are invited on all areas of WSD, and especially on :

* large-scale knowledge sources for WSD. Given the difficulty of building large-scale knowledge sources by hand, researchers have turned to "ready-made" resources such as machine-readable dictionaries and corpora. However, each of these kinds of resources poses problems, and none covers definitively the areas required to accomplish WSD.

* the problem of sense division. The disambiguation of word senses involves, a priori, determining what the appropriate sense distinctions for a given word are. Many studies have shown that the kinds of sense distinctions made by most everyday dictionaries (and even some computer dictionary resources such as WordNet) are too finely-grained, and in some cases not even appropriate, to serve the purposes of language analysis.

* combination of methods (statistical, rule-based) and knowledge sources (associative, collocational, phrasal, morphosyntactic, statistical, domain-related, etc.). It is now widely held that no single approach is complete enough for WSD, and that a combination of sources and methods is required. However, it remains to be determined how to most effectively combine methods and knowledge for WSD.

* assessment of the knowledge "needs" for WSD. There is no precise quantitative study on what the knowledge "needs" are in order to disambiguate a given word in a given context. For example, in the sentences, "I write on the page", "I spoke to the page", "The front page says..." it is obvious that different kinds of knowledge contribute more or less to the proper (and/or easy) disambiguation of "page" in the various contexts--i.e., the strong association between "write" and

"page", selectional restrictions on the verb "speak", and collocational information for "front" and "page", respectively.

All papers will be peer reviewed. Priority will be given to papers that have a strong empirical background and report precise, quantitative results.

Submissions: Hard copy submissions should be on letter-size paper (8.5 x 11 inches, or A4), double-spaced throughout, including footnotes and references. The paper should begin with an informative abstract of 150-250 words. Manuscripts must be written in English. Six copies should be sent to: Julia Hirschberg, CL Editor, AT&T Bell Laboratories, 600 Mountain Avenue, 2D-450, Murray Hill, NJ 07974, USA (Tel: +1 (908) 582-7496; Email: acl@research.att.com. Submissions in electronic form (LaTeX) must conform to the Computational Linguistics specifications, and be sent to <acl@research.att.com>, before April 1, 1996.

Machine Translation
Special Issue on New Tools for Human Translators
Guest editors:
Kenneth W. Church (AT&T Bell Labs) and Pierre Isabelle (CITI)

The Machine Translation journal is inviting submissions for a Special Issue on the theme of translation support tools.

Fifteen years ago, in his seminal paper "The Proper Place of Men and Machines in Language Translation," Martin Kay argued forcefully for what was then a radical new view of Machine Translation (MT), and has since become known as the "workstation approach" or the "tools approach." In contrast with fully-automatic MT and largely-automatic approaches such as Machine Assisted Translation (MAT) followed by post-editing, Kay advocated the more modest goal of building machines capable of **supporting human translators** in ways that conform to **their own normal work processes**. The machine would only be called upon to execute those portions of these processes that can be automated reliably. As our understanding of human translation progresses, the assistance will extend deeper and deeper towards its core aspects.

The obvious starting point for this programme was to provide translators with customized office automation systems: multilingual word processing, on-line dictionaries and term banks, text archiving and retrieval facilities, etc. But beyond that, where can we go? Short of full MT, what more can we do for translators?

Some promising concepts have been discussed in recent literature, such as:
* automated terminology extraction and management
* translation memories (including bilingual concordancing systems)
* translation checkers
* translation dictation systems

Systems based on some of these concepts have even begun to appear on the market.

We are inviting high-quality, original research papers relevant to this topic of novel tools for translators, including but by no means limited to the four areas mentioned above.

Submission Deadline: February 29, 1996; Notification Date: April 15, 1996; Camera ready copy due: May 31, 1996

Format: Please send hardcopy or electronic file (plain text or PostScript) to either guest editor. For final camera-ready versions, authors are invited to use the publisher's latex style file. The Kluwer.sty file can be obtained: 1) through an e-mail request to editdept@wkap.nl, mentioning the Machine Translation Journal; or 2) gophering to GOPHER.WKAP.NL; or 3) the World Wide Web at URL gopher://gopher.wkap.nl.

Guest Editors: Kenneth Ward Church, AT&T Bell Labs, office 2B-421, 600 Mountain Av., Murray Hill, NJ 07974, U.S.A. (Email: kwc@research.att.com; Fax: (908) 582-7550);

Pierre Isabelle, CITI, 1575 Chomedey Blvd., Laval, Quebec, Canada H7V 2X2 (Email: isabelle@citi.doc.ca; Fax: (514) 973-5757)

PUBLICATIONS RECEIVED

Journals

Computational Linguistics, *vol.21 no.3 (September 1995). -- vol.21 no.4 (December 1995).* Contents include reviews of "Machine translation" by D.Arnold et al. (Susanne Heizmann), "Compositional translation" by M.T.Rosetta (Bonnie J.Dorr), and "Speech-to-speech translation" by Hiroaki Kitano (Nigel Ward).

Elsnews *vol.4 no.5 (November 1995)*. Contents include: Resources: what is ELSNET's role? (Ulrich Heid). -- Applications-oriented research [on PaTrans] (Bente Maegaard). -- USA-German Verbmobil collaboration [interview with Dan Flickinger]. -- MorphoLogic bridging gaps between academia and industry in Hungary (Gábor Prószéky).-- *vol.4 no.6 (December 1995)*. Contents include: Looking back, looking forward (Steven Krauwer). -- The Text Encoding Initiative (Lou Burnard).

Language International, *vol.7 no.5 (October 1995). -- vol.7 no.6 (December 1995).* Contents include: Machine-friendly or reader-friendly? (William J.Niven) [on German-English MT.]

Language Matters: news and views from ALPNET, *September 1995.*

LISA Forum Newsletter, *vol.4 no.4 (November 1995)*. Contents: Localisation: beyond the software sector (Rose Lockwood). -- How do we improve project management skills throughout the localisation industry? (Gunnie Jacobsson). -- Reviewing the localisation project management course (Alex McDonnell). -- Change management in language engineering - ELRA (Robin Bonthrone). -- The fate of the localization industry and a call to action (Jaap van der Meer). -- The quest for improvement (Alain Linden). -- News from Ireland (Joc Sanders; Reinhard Schaeler). -- DOCWARE: the document localisation workbench.

Literary and Linguistic Computing *vol.10 no.3 (September 1995), vol.10 no.4 (November 1995)*

Logos Corporation Apprentice, *Spring 1995*

Machine Translation *vol.10 nos.1-2, 1995.* Special issue: Building lexicons for machine translation II, edited by Bonnie J.Dorr and Judith Klavans. Contents: pp.1-3: Introduction (Bonnie J.Dorr and Judith Klavans). -- pp.5-57: A lexicon for knowledge-based MT (Boyan Onyshkevich and Sergei Nirenburg). -- pp.59-92: Verb semantics for English-Chinese translation (Martha Palmer and Zhibiao Wu). -- pp.93-141: Towards a cross-linguistically valid classification of spatial prepositions (Arturo Trujillo). -- pp.143-184: Toward a lexicalized grammar for interlinguas (Clare Voss and Bonnie J.Dorr). *vol.10 no.3, 1995.* Contents: pp. 185-218: Combining corpus and machine-readable dictionary data for building bilingual lexicons (Judith Klavans and Evelyne Tzoukermann). -- pp.219-258: Automatic acquisition of semantic collocation from corpora (Satoshi Sekine and Jun-ichi Tsujii). -- pp.259-267: Book reviews.

Machine Translation Review (Natural Language Translation Specialist Group, British

Computer Society), *no.2 (October 1995).* Contents include: Lexical resources for MT: a survey (Adam Kilgarriff).

Multilingual Communications & Computing *no.10 (vol.7 issue 1), 1995.* Contents include: Accelerating learning with translation software [Stylus] (Vladimir Draitser). -- Software tools for professional translation (Colin Brace)

Natural Language Engineering *vol.1 part 2 (June 1995).* Contents: Several knowledge models and a blackboard memory for human-machine robust dialogues (Violaine Prince and Didier Pernel). -- Best parse parsing with Earley's and Inside algorithms on probabilistic RTN (Young S.Han and Key-Sun Choi). -- Poisson mixtures (Kenneth W.Church and William A.Gale). -- A rule-based phrase parser for real-time text-to-speech synthesis (Joan Bachenko et al.)

Books

Alan K.Melby with C.Terry Warner: **The possibility of language. A discussion of the nature of language, with implications for human and machine translation.** (Benjamins Translation Library, vol.14) Amsterdam/Philadelphia: John Benjamins Publ. Co. 1995. xxvi, 274pp. ISBN: 90-272-1614-2 [Review to appear in MTNI#14]

International who's who in translation & terminology. Traduction et terminologie répertoire biographique international. Union Latine/Praetorius Ltd./International Where+How/Infoterm, 1995. xxxviii, 426pp. ISBN: 0-9516572-59, 92-9122-002-7 [See also p. ?]

Conference proceedings

Second Language Engineering Convention: Language engineering, solutions for managers in business. Queen Elizabeth II Conference Centre, London, 16-18 October 1995. *Convention digest.* London: DTI, 1995.

Translating & the Computer 17. Papers from the Aslib conference held on 9th and 10th November 1995. London: Aslib, 1995. ISBN: 0-85142-351-5

Reports

Lockwood, Rose; Leston, Jean; and Lachal, Laurent: **Globalisation, creating new markets with translation technology.** London: Ovum Ltd., 1995. 413pp. ISBN: 1-898972-25-7 [see MTNI#12: 23-24]

Mason, Jane; Rinsche, Adriane: **Translation technology products.** London: Ovum Ltd., 1995. 424pp. ISBN: 1-898972-20-6 [See also p. ?]

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.