MT News International

Newsletter of the International Association for Machine Translation

Published in February, June and October

From the Editorial Board

Poll of MTNI Readers

MT News International is *your* newsmagazine. As John Hutchins told you in his February editorial, we would very much like for you, the members of IAMT and its regional associations, to provide us with feedback so that we can meet your needs and expectations to the fullest extent possible. We want to hear which features you like best, what types of coverage you would like to see added, and any other ideas you may have on how we can improve your publication.

To systematize the input, we are using a questionnaire format. We ask you to take a few minutes to respond to the questions in the questionnaire on the last page of this issue [?]. Most of them ask you to rate a statement or suggestion on a scale of 1 to 5. The best score is 5; it means that you strongly agree with the statement or are strongly in favor of the feature proposed. The lowest

score is 1 (not zero!). You may add comments on a separate sheet, keying each comment to the number of the corresponding question.

So, please take a look at this issue and the immediately preceding ones, sharpen your pencil, and fill in your responses. The completed form may be faxed to the Editor-in-Chief or your regional editor at the address given on page 2. An electronic version may be requested from eamt@cst.ku.dk or AMTAInfo@aol.com.

NEW SYSTEMS and PRODUCTS

Systran Receives $10 Million Contract
from the National Air Intelligence Center

[Press release May 1996]

Systran Software Inc. (SSI) has been tapped by the National Air Intelligence Center (NAIC) to develop machine translation software products and services for the intelligence community. The five-year contract, valued at approximately $10 million, includes the creation of software for translation from several Eastern European languages, including the first-ever Serbo-Croatian-into-English program.

The contract also covers maintenance and enhancement of software on nine existing language pairs, including Japanese, Chinese and Korean into English. Systran translation software is used at more than 30 sites within the intelligence community.

Under the terms of the contract, SSI will also continue development of its proprietary multilingual data retrieval software, which allows a user to query in English and receive findings in other languages.

The NAIC, headquartered at Wright-Patterson Air Force Base in Dayton, Ohio, has had a 28-year relationship with SSI. The company pioneered the development of a Russian-into-English MT program for the intelligence-gathering agency, which now includes more than half million words and operates at more than 90 percent accuracy on technical texts.

SSI has developed MT systems for 28 language pairs and recently brought its advanced mainframe technology to the PC with its SYSTRAN Professional program for Windows. The company also offers translation services for all languages and software development of new language pairs.

Globalink Announces New Internet Product:
Web Translator Software Translates Foreign Lannguage Web Sites

On March 5, 1996 Globalink announced its newest product: Web Translator. This new Windows software enables anyone surfing the Internet to translate sites currently shown in Spanish, French and German into English, or vice versa, by clicking a button on their Web browser software. Web Translator works with Netscape's Navigator 2.0. It is expected to have a street price of $49.95.

Web Translator's multilingual translation capability allows users to access foreign language Web sites and to produce draft translations in their preferred language. Translated pages maintain graphics, hotlinks and formatting. These translations can be created on-line, or alternatively, pages can be downloaded and saved to be translated and viewed off-line. The dictionaries for Web Translator include Internet terminology. Interestingly, the technology underlying Web Translator combines the latest MTAPI (Machine Translation Applications Programming Interface) with Globalink's existing Power Translator system.

Jim Lewis, Globalink's president commented, "We are very excited about Web Translator - it gives Globalink a fun and entertaining way to expose people to the endless possibilities of translation technology. Web Translator is the only multilingual product that actually translates foreign language

Web sites from around the globe.... People of all ages will enjoy using this product to expand their horizons and see the true scope of the Internet."

Globalink's own Web site at http://www.globalink.com maintains an updated list of foreign language sites of general interest for use with Web Translator. Web Translator 1.0 can currently be used with Netscape's Navigator 2.0. Support for other browsers, including Mosaic, will soon follow and will be available free for registered users to download from the Globalink Web site. Depending on the user's PC, Web Translator can be run in either 16-bit or 32-bit mode on either Microsoft(Windows 95 or Microsoft Windows 3.1. The product comes on a single CD-ROM that includes both domestic and localized versions. The product will be available through most major retail computer stores including COMP USA and Computer City.

News from Toltran: Patent Granted

[Press release, May 1996]

Toltran Ltd. of Barrington, Illinois has just been granted United States Patent No. 5,490,061. This patent covers advanced technology and is the first to use the concept of Chaos in a computational linguistics setting.

The patent is particularly important now due to the ever-expanding use of search engines on the Internet, and has numerous applications in the fields of data warehousing, database searching and the related problem of language translation. The database match or query problem widely encountered in language translation lookups is just a small subset of the general search problem in English or any language, and this was used as the demonstration case. The sophisticated morphological analysis is to be considered a breakthrough in its field, and the search engine algorithms are adaptable to almost every conceivable type of use, such as worldwide web browsers, real-time satellite transmission translations, OCR, speech recognition systems and natural-language-to-computer-language CASE tools.

The technology is an offshoot of Toltran's previously-issued patent no. 4,864,503, which introduces previously unattained accuracy in translation via its object-oriented modular design, in which languages are considered source or destination modules and all feed through a hub-and-spoke system. This allows code reuse, along with greatly reduced database and program size. The inventor, a computational linguist, polyglot and systems engineer, was listed in several editions of "Who's Who In America". The company with its intellectual property porfolio is currently seeking an acquisition scenario.

The patent is entitled, "Improved Translation System Utilizing A Morphological Stripping Process To Reduce Words To Their Root Configuration To Produce Reduction Of Database Size". A more detailed company and technology profile may be found at this website: http://www.knowledge.co.uk/xxx/mpcdir/products/ts_002.htm

Please reply to Stanley Tolin (Tel: +1-941-387-0195; Fax: +1-941-387-0235; Email: stolin@mail.flnet.

PRODUCTS and SYSTEMS

PENPARS and "Translation Office"
New Products by Lingvistica '93 Co. and ETS Ltd.

Michael Blekhman

MT systems in the PEN environment

Being a linguist and translator, I know but well that post-editing a "raw" translation generated by a computer program is not an easy task, especially if you want to prepare a publication quality

document. There are several typical operations the post-editor has to perform dozens if not hundreds of times when "polishing" the machine-produced text, transposing words being one of the most tiresome ones.

My personal experience shows that using powerful general-purpose text processors such as WinWord is rather hard; what is really needed is a special, translator-oriented editing tool.

Lingvistica '93 Co. (Kharkov, Ukraine) united efforts with a Kiev-based team of programmers led by Dr.Alexander Kazakov, who has great experience of working on neuron computers, to develop what we call the PENPARS package. It combines three MT systems with a powerful pen editor. The systems are: PARS (English-Russian-English), PARS/U (English-Ukrainian-English), RUMP (Russian-Ukrainian-Russian). The user may activate any of the above MT systems directly from PenEdit, as well as post-edit the raw translation.

The major advantage of the editor is the possibility to work with the images created by means of the pen as if they were standard objects, and these images may be connected directly with the operations they correspond to. This eliminates the main restriction characteristic of the computer keyboard, i.e. a limited number of keys and their combinations a human being can master.

PENPARS supports a multi-window operation mode, and, what is especially useful for the translator, the latter may process the text in the natural way, including all kinds of operations with characters, tables, and text blocks. For instance, to eliminate a character, you simply cross it out as if it were written on a sheet of paper. To move a word or a phrase to another position in the sentence, you underline it with the pen and mark the place to insert the word. Everything is very convenient simply because it's quite natural.

PARS & Polyglossum project

Lingvistica '93 Co., ETS Ltd (Moscow, Russia) and two other Russian firms have issued several compact disks of the "Translation Office" series. I don't claim that the name is the best one because a translation office presupposes much more than we can offer. At the same time, experience shows that these products can be of use for those numerous people who deal with English and Russian scientific, business, and technical texts.

The disks contain four systems, in various combinations of programs and dictionaries: PARS for Windows and DOS; Polyglossum for Windows and DOS, a dictionary-support tool with the world's largest English-Russian-English dictionaries; German-Russian bi-directional dictionaries on CD-ROM are also available; AutoR, a character recognition program for DOS, by the OCRUS company; EnglishGold, a Windows-based multimedia English teaching system by the Multimedia Technologies company.

The products has proved to be rather popular in Russia: more than 1,000 copies have been sold within two months. This also shows that machine translation is a tool for "broad masses" of people in former Soviet Union.

ThoughtTreasure for English-French Translation

[Press Release, May 1996]
Erik Mueller has released the ThoughtTreasure program for English/French machine translation under the terms of the GNU General Public License. The program and associated book are available on the web at: http://www.panix.com/~erik/tt/html/tt.html

ThoughtTreasure consists of:
. English and French lexicon - 50,133 words/phrases.
. Ontology/database - 21,521 concepts, 37,162 assertions, concise format, space represented by 2-d grids connected by wormholes, grids for house/restaurant/theater/street/subway.
. 67,966 lines of ANSI C source code.
. Syntactic parser - base component, filters, transformations.
. Surfacy semantic parser - intension/extension, relative clauses, appositives, genitives, tense/aspect.

. Anaphoric parser - deixis, determiners, pronouns, c-command.

. English and French generator.

. Text agents – part-of-speech tagging, names, dates, phone numbers, products, email headers.

. Planning agents - simulating human and device behavior, graspers, containers, ptrans++, atrans++, mtrans++, devices (telephone/TV/car), interpersonal relations.

. Understanding agents - converting surface parse into detailed understanding, steering planning agents, contexts, common sense knowledge, emotions, goals, question answering, asking clarifying questions, appointments, sleep, grids by analogy.

. Learning - extracting data from tables/Usenet articles, learning new words using derivational rules, learning new inflections by analogy.

. Tools/utilities - shell, concordance, report generation.

. Applications - translation, information retrieval, conversation, calendar maintenance, emotion understanding, story understanding, story generation, daydreaming, dictionary generation, linguistic studies and indexing.

ALEP-3.1 Released

[Extract from ELSNET]

Cray Systems announce the release of the Advanced Language Engineering Platform (ALEP) Version 3.1. ALEP-3 is the Quintus Prolog version of ALEP-2. ALEP-3 is distributed on QIC-150 tape with a printed installation guide. ALEP documentation and related reports are supplied as part of the standard ALEP-3.1 distribution. Info files are also supplied for online browsing using the XMINFO graphical viewing tool.

Software Requirements for ALEP-3.1: GNU Emacs 19, OSF/Motif 1.2, a Quintus Prolog 3.1.2. development licence is required for building the system. ALEP is available on Sparc with operating system Solaris 2.x or SunOS 4.1.x (= Solaris 1). Hardware requirements: at least 80 MB swap space; Minimum 16 MB core memory.

Send requests for ALEP-3.1 Distribution Tape to: ALEP Support, Cray Systems, 151 rue des Muguets, L-2167 Luxembourg (Tel: +352 42 77 44; Fax: +352 43 95 94; E-mail: alep-support@cray-systems.lu)

ALEP courses

The Institute of Applied Information Science (IAI) is planning to conduct ALEP courses (introductory courses and advanced user courses) and an ALEP workshop during this year. The courses and the workshop will not be supported by the European Commission (as this was the case for previous events). Therefore, a participation fee of approximately 100 Ecu will be necessary. For more information on ALEP, AUG, etc. please visit the ALEP Web page at URL http://www.iai.uni-sb.de/alep.

JICST's MT Service on the Internet

[Extracted from LINGUIST List (vol-7-550) posting by Patti O'Neill-Brown, Japan Technology Analyst, U.S. Department of Commerce]

A machine translation system, developed especially for translating scientific and technical Japanese information into English is now available via Internet. A Macintosh or IBM-compatible PC with a bilingual OS (such as Windows 3.1J) is required.

The Japanese-English Machine Translation Network System has been developed by the Japan Information Center of Science and Technology (JICST), a public corporation established by the Japanese government in 1957 as Japan's key center for scientific and technical information

services. JICST's machine translation system, an improved version of a system originally developed in the 1980s through a joint project between Kyoto University and the Science and Technology Agency, was initially designed for internal use to produce English-language S&T databases from Japanese-language databases. In January 1996, JICST decided to make the translation system available via Internet.

How to use the JICST translation service:

     1. Enter JICST's "Translation Network Service" through JICST's World Wide Web Homepage (http://www-jmt.jicst.go.jp/index-E.html).

     2. Send the original Japanese document to JICST in text format utilizing a mail software such as Eudora. The Translation Network System processes the translation and sends the English output back to the customer's e-mail address.

     3. Open mailbox and download the English text.

Translation Fees:

One yen per character of the English translation output (plus 3% Japanese consumption tax.)

Initial Registration Requirements:

To use the translation service for the first time (including trial use), submit a written application, via FAX, to JICST. To obtain an application form and instructions, those who live outside Japan should contact: JICST Overseas Service Division (Fax: +81-3-5214-8413) Individuals residing in Japan should contact the closest JICST branch office (see contact addresses at the end of this report).

Limitations:

     1. The original document must be less than 20,000 Japanese characters in total length for each translation task. In other words, longer documents must be divided into segments of less than 20,000 characters each.

     2. Only JIS Kanji code can be used.

     3. Since translations are performed only by machine with no pre-editing, post-editing or re-writing by human operators, users of the system must be prepared to carry out these functions themselves.

JICST announce PC-Based MT System

JICST has also developed their machine translation system for use on a Macintosh or PC. They announced on March 1, 1996 that a CD-ROM package of their machine translation program, combined with a dictionary of about 400,000 S&T terms, will be put on sale this summer for about $500. Although several commercial PC-based Japanese-English translation systems are available, most are designed for general office or daily use and have relatively small dictionaries. In terms of translation speed and quality, JICST's system is said to be about the same as many commercial systems. Thus, the JICST system may offer advantages for the translation of scientific and technical information.

     Hardware requirements: Macintosh or Power Macintosh (running on bilingual System 7 or System 7.5 environment), or IBM-Compatible PCs with Windows 3.1J or Windows 95 OS; Hard Disk: 300 MB or more; Memory: 16 MB or more; A CD-ROM drive connected to the computer.

     The system will include a dictionary of 400,000 scientific and technical terms. JICST is also planning an auxiliary dictionary of about 200,000 medical terms (price yet to be decided). In addition, users can develop their own "user dictionary" (a feature not available for the Translation Network System on the Internet).

     For further information see JICST's Home Page at: http://www-jmt.jicst.go.jp/index-E.html or: Fax to JICST Overseas Service Division (+81-3-5214-8413)

Four Windows software products from NEC

[Extracts and summaries from AAMT Journal nos.12 and 13] – *John Hutchins*

Mr.EJ Dic Reader allows easy consultation of a dictionary on a PC. English language files are read into *Mr EJ Dic Reader*'s own window; a click on a word brings up Japanese counterparts, which with the "Paste equivalent" function can then be pasted below the original. "Batch consultation" allows all words in a file to be looked up en bloc, and "Write Memo" allows insertion of Japanese translations as memos.

Mr.EJ Dic Reader comes with *Ms.EJ Dic Scanner*, which provides consultation of the internal dictionary from inside such programs as an Internet browser, communications software and word processor running on Windows. The standard dictionary contains 90,000 words, and optional technical dictionaries totalling some 600,000 terms. User dictionaries have no limits.

An extension facility of *Mr.EJ Dic Reader* is a machine translation interface, allowing it to be used as an MT system if used with *E-J MT Adaptor for Internet* (see below) or with the *j*London/EJ* system from Kodensha. Another extension is a CD-ROM scanning interface, enabling Mr.EJ Dic Reader to run in conjunction with other commercial software for CD-ROM data retrieval, e.g. to consult a CD-ROM dictionary.

The basic system costs 9,800 yen; it requires a PC 486 with minimum 20MHz, 8MB RAM, 18MB free disk space and Windows 3.1 (Japanese version).

The *E-J MT Adaptor for Internet* combines with an Internet browser, communications software and word processor to do English-to-Japanese translation.

Translation is activated by mouse, either of full sentences or of individual words; translation style can be specified. The standard dictionary contains about 90,000 words, with the option of adding more entries. It comes at 14,800 yen, requiring PC-9821 or DOS/V with CPU 486, 66MHz or higher, 16MB RAM, 68MB disk space, and Japanese Windows 3.1.

Netscape/ej is a product which provides *Netscape Navigator* with facilities for consulting an English dictionary. Japanese equivalents of English words can be seen during WWW browsing. The internal dictionary contains 12,000 words. If combined with *E-J MT Adaptor for Internet* it can be used as a Netscape Navigator capable of English-to-Japanese translation. The list price is 6,000 yen. The software requires PC-9821 or DOS/V, with CPU 486, 20MHz, 8MB RAM, 7MB disk space, and Japanese Windows 3.1.

NEC's latest product is the *Japanese-English Dictionary Referencer* software to help Japanese to write English. It comes pre-installed in the NEC's personal computer "CanBe", and also packaged in "Lunch Pack Ver.2" (for Windows 3.1: 9,800 yen), and in "Youthful Pack" (for Windows 95: 3,800 yen)

The *Japanese-English Dictionary Referencer* is similar to the Kana-Kanji conversion interface used on Japanese wordprocessors in that the operating procedure for Kana-Kanji conversion is extended to Japanese-English word conversion. When a Japanese word is keyed in, the Japanese-English dictionary (containing about 100,000 entries) finds the corresponding words and displays them. Users then can select the appropriate word with the cursor. To assist word selection the system provides an idiom dictionary (ca.10,000 entries) with auxiliary information on usage. In addition the Japanese-English Dictionary Referencer enables searches in any commercially available EP-WING Standard compliant CD-ROM English-Japanese and Japanese-English Dictionary.

The software can be used with any wordprocessor or editor, or e-mail software. It can be activated from any software running on Windows 3.1 or Windows 95. System requirements are: PC-9800 Series, Japanese Windows 3.1 or Windows 95, CPU 486, 20 MHz or more, 8 MB RAM, 10MB disk space. For more information contact: NEC Personal Computer Information Center, Aqua City Shibaura Bldg. 4-16-23 Shibaura, Minato-ku, Tokyo, 108, Japan (Tel: +81-3 3452 8000)

The Translator-Mini Ver.1.5
for Macintosh from Catena Corporation

Catena released its new English-to-Japanese translation software for Macintosh, *The Translator-Mini Ver.1.5*, on the 14th September 1995. Designed for simple translation work and claimed to be an easy-to-use high-speed system, the latest version has an enriched dictionary and improved grammatical processing for more accurate translation. The makers report sales of 10,000 packages since release.

The standard dictionary contains 67,000 words, the package includes a 32,000 dictionary of computer and basic science terms. There are also optional technical dictionaries (price 30,000 yen upwards) for medicine, mechanical engineering, electrical engineering, geophysics, etc. Translation is fastest on a PowerMac running Kanji Talk (version 7.1 or up) in the PowerPC Native mode.

The product is sold at 14,800 yen. Minimum requirements are Macintosh with Kanji Talk 7.1, CPU M68030, 8MB RAM, 16MB disk space. For further information contact: Software Products Division, Katena Corp., 1-15-2, Ochiai, Tama City, Tokyo 206 Japan (Tel: +81-423-38-1118)

PENSEE for Internet:
Windows 95 and Windows NT versions
from Oki Software Inc/Oki Electronic Inc.

PENSEE for Internet is a WWW translation software that runs on Windows 95 or Windows NT. It is a modified and expanded version of *PENSEE for Internet Server Model* which was the first Japanese software to translate WWW server information on line. It operates in conjunction with any browser (such as NetScape, Explorer, AirMosaic, Ichitaro's JustView) with no special procedures.

Translation does not affect the screen layout format. The Japanese translation looks the same on the screen as the original English as they come in the identical format. The makers claim a translation speed of about 10 sec for a typical WWW home page, and about 20 sec for a very long page (if a PC with Pentium 100 MHz or its equivalent is used).

PENSEE for Internet comes with a dictionary for Internet, and optional special dictionaries are available, listing as many as 150,000 words and covering 9 different specialties. These dictionaries can be used with the special dictionaries for the Oki Software's professional translation software *PENSEE V*.

The product sells for 9,8000 yen, special dictionaries are 28,000 yen (fields covered: business, economics, basic technologies, information and communication, automobile engineering, science, medicine, machines.) It requires a PC with Windows 95 or Windows NT, and capable of accessing WWW; minimum 8MB RAM, 30MB disk space, Modem, and browser (NetScape, Explorer, etc.) Installation is on CD-Rom. For more information contact: 2nd Sales Section, Sales Department, Oki Software Inc. (Tel: +81-3 5970 9952, Fax: +81-3 59 70 9902)

Rupo JW-V700
English Language Translation
Supporting Capability from Toshiba Inc.

The *Rupo JW-V700* is the first personal wordprocessor capable of English-to-Japanese translation. It comes with built-in OCR. Printed English sentences can be read into the wordprocessor where they

are directly translated. English sentences keyed in manually are spell-checked before translation. The results can be printed on the right half of each page with the original printed on the left half, or printed with alternating English and Japanese lines. There are two modes: "Batch translation" for whole texts, and "Consecutive translation" for one word at a time, enabling on-screen correction.

The standard dictionary contains approximately 61,000 words. About new 2,300 words can be entered in the user dictionary; help for users with grammatical information is provided.

The JW-V700 is priced at 250,000 yen (without tax). For more information contact: Wordprocessor Personal Information Equipment Department (Tel: +81-3 3457 2903)

Dr. SURF for Windows
English-Japanese translation support software
from Media Vision Inc.,
Kyushu Matsushita Electric R & D Center

On Nov. 22, 1995, Media Vision, under the license from Kyushu Matsushita Electric, released an English-Japanese translation support software *Dr. SURF for Windows*, capable of translating both English text and Internet WWW browsers.

Dr. SURF operates in 3 translation modes: "Text Translation Mode", "Real-time Translation Mode" for Internet, and "Cache Translation Mode". The original English and its translation are displayed on alternate lines on the same browser. Translation proceeds phrase by phrase and in real-time.

The "Cache Translation Mode" allows data disk-cached on NetScape Navigator 1.X to be translated off-line. And *Dr. SURF* can be used also for translating English text on a dialogue-type editor. The product comes with a 70,000-word dictionary, a user dictionary that can register up to 50,000 words, and a 30,000-word dictionary for Internet. 18 different special dictionaries are optionally available.

The system requires: Japanese Windows 3.1, or Japanese Windows 95, 8MB minimum RAM, 30 MB disk space, CD-ROM drive, compatible TCP/IP (Chameleon, Trumpet Winsock for Windows 3.1), WWW Browser: NetScape Navigator 1.X recommended. Standard retail price is 14,800 yen (without tax). For more information contact: Media Vision, 3F Kudan Eyelex Bldg., 1-14-21 Kudankita Chiyoda-ku, Tokyo, 102, Japan (Tel: +81-3 3222 3908)

USERS and APPLICATIONS

Current applications in Europe

Paul Hearn

With European integration moving ever closer, it is easy to pause with wonder at the increasing amount of information that travels from one corner of Europe to another every working day, encoded in words and phrases in numerous languages. With its multilingual heritage, the European Union must rank as *the* no. 1 market for machine translation products and services.

However, do current machine translation products offer sufficient robustness to meet the needs of Europe's multilingual information society? And importantly, to what extent are companies and organisations in the European Union building home-grown expertise in machine translation? These questions are not superfluous; they are vital to the positioning of Europe, not only as the no. 1 *market* for machine translation technologies, but also as the *world leader* in language engineering applications. Also, and not less importantly, European expertise in machine translation will help bring about the successful and smooth transition to a European multilingual information society.

This column focuses on a number of representative activities occurring at present in Europe in the machine translation area. In this article, we will take a brief look at one of Europe's leading machine translation users, the Germany-based *SAP AG* company, which is using machine translation now to meet its huge internal documentation needs. We will then consider some recently-concluded European projects, and look at some of the users hoping to exploit the technologies developed in the framework of these projects.

SAP must rank as one of Europe's major users of machine translation, having translated automatically a total of 8 million words this year alone, and 20 million words since 1991. The German-based company uses the US/European-developed METAL system for German-English translation, and the US-developed LOGOS system for English-French translations. Together these systems help to translate the technical documentation, user manuals, customer error notes and user implementation guides which support SAP's product range. Using a layered post-editing system, with three quality levels, SAP achieves productivity which, according to Jennifer Brundage of SAP's MT Group in Walldorf, is "2 to 3 times higher than pure human translation". The MT Group at SAP doubled in size last year, and still finds it difficult to meet the demand for potential automatic translation of the *4 million lines* of documentation which form SAP's total product literature.

SAP's machine translation efforts currently focus exclusively on meeting internal needs. However, the company may consider marketing its machine translation expertise to the outside world at some stage in the future. The LE1-funded OTELO project, of which *SAP* is a member, along with *Lotus Development Ireland,* representatives of commercial machine translation systems (*CST Denmark, GMS Germany* and *Logos Germany*), and a number of associated users and providers of human and machine translation services, aims to design and develop a comprehensive automated translator's environment, combining in a single interface disparate natural language processing (NLP) programs including machine translation, translation memory, and an array of automated translation and editing tools. Importantly, OTELO aims to allow many new potential users to use machine translation for the first time, by offering access over public and private networks and the world wide web. OTELO may yet provide the environment and impetus needed to bring SAP's machine translation experience to the wider world.

Also funded by the European Commission, only this time in the context of the MLAP93 program, TRADE (Jan 1994 - May 1995) addressed the specific subdomain of translation of social security reports and statistics, within the context of the Italian *Istituto Nazionale della Previdenza Sociale* (INPS). As a bonus, it was also intended that the TRADE project would deliver tools which could also be used by software companies with localisation challenges. The machine translation demonstrator developed during the project included complete sets of translation pairs for English, Spanish and Italian, based in part on an inherited Eurotra formalism. According to Joan Manel Espejo, Project Manager of TRADE and Responsible for European Programmes at *CCS,* a Spanish software developer, the demonstrator has been tested both at *INPS* and at *CCS*. In the Spanish context, it is now used daily and has proved to be "a helpful tool in the translation process for several *CCS* manuals since November 1995". According to Espejo, the system is used mainly to provide pre-translations of documents: "*CCS* translation experts estimate that this pre-translation approach increases their translation throughput by 30%". As a result of this initial success a pilot machine translation system has been started at *CCS* for automatic translation of a range of business management software and supporting documentation. At *INPS* the use of the demonstrator has been limited due to the need to improve translation quality by adding new terms to a growing electronic dictionary of social security terms. A total of 10,000 - 11,000 entries are now available in English, Italian and Spanish. *INPS* is optimistic that the demonstrator may go into production in the near future, and help to translate some of the huge number of social security documents produced by the approximately 200 *INPS* subsidiaries throughout Italy.

While not addressing machine translation per se, the LRE2-funded SECC project (A Simplified English Grammar and Style Checker/Corrector) covered the important area of controlled language checking. The project included partners from *Siemens Nixdorf, Belgium, Alcatel Telecom, Belgium, Cap Gemini Innovation, France, University of Leuven, Belgium,* and *Sietec, Germany*, and

is scheduled to end around the time this article is published. SECC has resulted in the development of a tool to help technical authors to write according to the rules of simplified English. An important additional feature is that it also helps Dutch, French and German-speaking authors when producing English technical texts, by flagging common mistakes due to language interference. The value of such tools should not be underestimated in Europe. The SECC tool may also be used as a pre-processor of English for machine translation. While SECC is a wholly European development, the project initially benefited from significant input from consultants from Boeing in the US. A first version checker is now available as a marketable product and is currently being used at *Alcatel Telecom* in Belgium and by *General Motors* in the US. Commercialisation rights have been transferred to *LANT*, of Heverlee, Belgium, a company specialising in exploiting machine translation products and technologies, and *Cap Gemini Innovation,* so we can expect to see more of SECC in the future.

The success of SECC goes beyond the immediate commercialisation possibilities, since it has managed to generate considerable interest in controlled language applications in general, by playing a catalyst role in European RTD related to controlled languages. Dr Geert Adriaens, NLP Project Manager at Siemens Nixdorf's Software Centre in Liège, Belgium, and Project Manager for SECC, sees the longer term view made possible by projects such as SECC: "Attempts to control or streamline language in technical documentation could lead to the definition of standards for document quality which go far beyond character sets or general document layout conventions. If we succeeded in creating an ISO standard taking over a core set of rules and vocabulary from the different controlled language definitions existing to date, we would not only have proved the importance of our field to the real world out there, but we would have added an important building block in the construction of an overall product and service quality system."

Developments at the European Commission

Dorothy Senez

Business as usual.....

The Context

The Systran machine translation system, developed by the Commission since 1976, can produce 2 000 pages of raw translation per hour. Thanks to internal electronic mail, any of our officials in Brussels or Luxembourg can have a translation back in a few minutes. A total of 16 language pairs are available and their quality depends on the stage of their development. The most used, and the most reliable, are French into English and English into French. Less progress has been made with the Germanic languages.

Use of MT

A sharp increase in the use of MT in recent years has been due, not only to promotion, but also to the widespread adoption of electronic mail. Only 20,000 pages were produced in 1990; in 1995 170,000.

Where Systran dictionaries have been programmed to provide specialized terminology for specific, repetitive text types, some in-house translators, particularly the Spanish, are using MT to produce final translations of high quality. However, 70% of users are non-linguists working in the Commission's administrative departments. They use EC Systran in three ways. First, MT provides a rapid means of revealing the subject and gist of a document in a language unknown to the reader. Secondly, it can aid the drafting of documents (a French-speaking official who needs to produce an urgent internal memo in English may prefer to draft the paper in the mother tongue and then have the MT output corrected). Thirdly, it is used for the rapid translation of, say, minutes or technical reports.

Most of the machine output is edited in the requesting department, either by the requester himself or by colleagues who are fluent in the target language. Nevertheless, users are also able to call on a rapid post-editing service for papers that are not destined for publication (such as working documents and internal notes). This service, which relies on a network of freelance translators, helps keep a tighter linguistic check on urgent texts.

Linguistic Development

A distinction has been drawn between targeted development (the introduction of specific terminology requested by users) and basic linguistic development (general improvements, whether grammatical, syntactical or lexical). Every three months, a minimum of 9,000 new "low-risk" entries and 500 more complex improvements are introduced into the system. Close co-operation between volunteer in-house linguists and the system's developers have produced excellent results. Moreover, a large number of entries from the Eurodicautom terminology data base have been imported into the Systran dictionaries and this has involved a certain amount of consolidation work.

Euramis interface

The European Advanced Multilingual Information System (Euramis) refers to a windows interface which is currently used in the Commission's Translation Service. It has been designed to facilitate access to a number of services (machine translation, automatic look-up of the titles of legislative documents from the Celex data base and automatic terminology look-up). After final testing, the interface is due to be distributed to the other Commission departments, when a further increase in the use of MT is expected.

And some new ideas......

Euramis II Project

The aim of the Euramis II project is to achieve a vast multilingual memory. The Translation Service produces more than a million pages of translation every year, and although PCs are becoming increasingly powerful in terms of both storage capacity and processing, it will not be possible to store all potentially reusable resources locally, at least not in the foreseeable future. All the linguistic data of the Translation Service (terminology, phraseology and translation memories) will be gathered together in multilingually aligned texts (MATs) in a central base known as the Linguistic Resources Database (LRD) and accessed by powerful search tools. The current Euramis interface, which is serving as a test bed, will eventually become a doorway leading to a variety of applications in the field of natural language processing. Ultimately, the Commission's MT system will be integrated with the whole gamut of linguistic tools currently under development.

Systran Dictionaries

With an eye to these future developments, work is on hand on several aspects of the Systran dictionaries.

Exporting of Systran Dictionaries

It was decided, to ensure compatibility with the Commission's LRD, to convert the Systran dictionaries into a Terminology Interchange Format (TIF) under Unicode. As data transfer such as this was not anticipated by the system's originators, it has proved to be a complex project and further work will be required to complete the task.

Simplified Coding

The aim of this project is to enable users to code Systran dictionary entries themselves. Terms coded in this way will be stocked in a user dictionary for immediate use. The advantage is that users will no longer have to wait to have their terminology introduced centrally by the development team. Such a tool does imply, however, that users are prepared to assume a more active role in the production of the MT output.

Dictionary Structuring

A direct consequence of the simplified coding project is the creation of targeted dictionaries compiled for the specific needs of users or groups of users (individuals, translation units, working groups or even entire directorates). A number of questions have to be clarified by the development team: how to integrate these specialized dictionaries with the main Systran dictionaries; how to make them available to other MT users at the Commission; and to what extent user feedback of this kind can be exploited and reintegrated into the main system.

Access To Systran Dictionaries

We try to improve not only the service to users, but also the tools we make available to those involved in developing the system. Linguists working on improving the quality of the various

language pairs can now consult the Systran dictionaries from their PCs, relaunch translations, and search for particular instances of words or expressions by means of KWICs (Key Word in Context).

Evaluation

A concept recently introduced has been that of quarterly evaluations of the EC Systran system to quantify progress made, on the basis of an empirical methodology. Targets were set for the net percentages of improvement to be attained per language pair, depending on their stage of development.

However, changes on the horizon......

From an Experimental to an Operational Environment

For the past twenty years funding for the linguistic and technical development of the EC Systran translation system has been provided in a research context by DG XIII (Directorate-General for Telecommunications, Information Market and Exploitation of Research) under the Multilingual Action Plans (MLAPs). The Translation Service has played an active and joint role in matters relating to language. Faced with a steady annual increase of 20% in the use of MT over the past 5 years, and the fact that, for certain language pairs at least, MT has become an operational issue, the Commission has had to review its policy. Since DG XIII is now concentrating its efforts on its new MLIS (Multilingual Information Society) programme for promoting the linguistic diversity of the Community in the information society, funding for the institution's MT needs must be sought elsewhere. The Translation Service is currently looking at ways of taking over the project.

Taking Stock

A crossroads has been reached in the history of MT at the Commission and it is now imperative to decide which way will lead to the best and most economical use of the facility. Three studies are being prepared:

An in-house study, in close collaboration with the users themselves, to determine users' needs in regard to speed and quality.

A technological and market study of MT providers and users, and of MT systems and services and their costs. This will provide a clear picture of the state of the whole of the MT market.

A cost/benefit analysis to be based on the results of the two previous points. On the strength of this, decisions will be taken regarding the future orientation of MT at the Commission.

The Future

There can be little doubt that MT at the Commission is here to stay; but it is just as sure that the context is bound to change. One of the problems that must be faced sooner or later will be how to maintain a system on an IBM compatible. At some point it will be necessary to move from the current operating system to a more open one. The service will have to be provided from different budgetary resources. This means internal reorganization, an evaluation of the legal context, and calls for tender for MT systems and services. Although much upheaval lies ahead, a core user population of over 2000 users cannot be ignored and we must continue a strenuous search to find the best way of exploiting the resources at our disposal in a spirit of quiet optimism.

CORPORA and SERVICES

New Releases from Linguistic Data Consortium

[Extracts from LINGUIST]

Cellular TIMIT Speech Corpus (CTIMIT)

The CTIMIT corpus is a cellular-bandwidth adjunct to the TIMIT Acoustic Phonetic Continuous Speech Corpus (NIST Speech Disc CD1-1.1/NTIS Pb91-505065, October 1990). The corpus was contributed by Lockheed-Martin Sanders to the LDC for distribution on CD-ROM media.

Radio Broadcast News: Continuous Speech Recognition Corpus (Hub-4)

This set of CD-ROMs contains all of the speech data provided to sites participating in the DARPA CSR November 1995 Hub-4 (Radio) Broadcast News tests. The data consists of digitized waveforms of MarketPlace (tm) business news radio shows provided by KUSC through an agreement with the Linguistic Data Consortium, and detailed transcriptions of those broadcasts. The software NIST used to process and score the output of the test systems is also included.

Spanish News Text Collection

The Spanish News Corpus consists of journalistic text data from one newspaper (El Norte, Mexico) and from the Spanish-language services of three newswire sources: Agence France Presse, Associated Press Worldstream, and Reuters. (The Reuters collection comprises two distinct services: Reuters Spanish Language News Service and Reuters Latin American Business Report.) The period covered by these collections runs from December 1993 (for APWS and Reuters) or May 1994 (APWS) through December 1995.

Further information about the LDC and its available corpora can be accessed on the Linguistic Data Consortium WWW Home Page at URL http://www.cis.upenn.edu/~ldc. Information is also available via ftp at ftp.cis.upenn.edu under pub/ldc; for ftp access, please use "anonymous" as your login name, and give your email address when asked for password.

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

---

CONFERENCE REPORTS

Conferences in Leuven and Dublin

Two conferences in March and May included aspects of MT in their programmes.

The first (26-27 March) was **CLAW 96**, an international workshop on Controlled Language Applications, organised under the chairmanship of Geert Adriaens by the Centre for Computational Linguistics at the University of Leuven, Belgium. While nearly all papers will have been of some interest in MT circles since so much research is focussed on the use of controlled languages and sublanguages in the field, some papers can be highlighted as being specifically concerned with MT as such:

Controlled language correction and translation (Pim van der Eijk, Michiel de Koning and Gert van der Steen) [Cap Volmac]

Controlled language and knowledge-based machine translation (Eric H.Nyberg and Teruko Mitamura) [Carnegie Mellon University]

Controlled English advantages for translated and original English documents (Phil Hayes, Steve Maxwell and Linda Schmandt) [Carnegie Group, and Caterpillar Corporation]

Controlled language support for Perkins Approved Clear English (Shona Douglas and Matthew Hurst)

The Controlled Automotive Service Language project (Linda Means and Kurt Godden) [General Motors and METAL]

The proceedings are available from: Centre for Computational Linguistics, Katholieke Universiteit Leuven, Maria-Theresiastraat 21, Leuven, Belgium (Tel: +32-16-325088; Fax: +32-16-325098; Email: claw96@ccl.kuleuven.ac.be; WWW: http://www.ccl.kuleuven.ac.be/claw/)

The second was held 9-11 May at Dublin City University. This was entitled **Unity in Diversity? International Translation Studies Conference**. The conference included three sessions on Computer-aided translation and interpreting. Among the presentations were:

The problem with machine translation (Reinhard Schäler)

Practical experience of CAT tools in the software localisation industry (Sharon O'Brien)

Using Logos: the Corel experience (Orlagh Neary)

Machine translation as a model of human translation (Paul Bennett)

Technology impact assessment: translation studies and its contribution to MT evaluation (Nicole Klingenberg)

A relevance-based approach to automatic interpreting (Birte Schmitz) [on Verbmobil]

Towards computer-aided interpreting: state of the art, limitations and future directions (Dieter Huber) [on ATR project]

Abstracts of presentations were prepared for the conference. It is intended to publish the full proceedings later. Information: Jennifer Pearson, SALIS, Dublin City University, Dublin 9, Ireland (Email: pearsonj@dcu.ie)

From the Archives...

ALPAC: the (in)famous report

John Hutchins

The best known event in the history of machine translation is without doubt the publication thirty years ago in November 1966 of the report by the Automatic Language Processing Advisory Committee (ALPAC 1966). Its effect was to bring to an end the substantial funding of MT research in the United States for some twenty years. More significantly, perhaps, was the clear message to the general public and the rest of the scientific community that MT was hopeless. For years afterwards, an interest in MT was something to keep quiet about; it was almost shameful. To this day, the 'failure' of MT is still repeated by many as an indisputable fact.

The impact of ALPAC is undeniable. Such was the notoriety of its report that from time to time in the next decades researchers would discuss among themselves whether "another ALPAC" might not be inflicted upon MT. At the 1984 ACL conference, for example, Margaret King (1984) introduced a panel session devoted to considering this very possibility. A few years later, the Japanese produced a report (JEIDA 1989) surveying the current situation in their country under the title: *A Japanese view of machine translation in light of the considerations and recommendations reported by ALPAC.*

While the fame or notoriety of ALPAC is familiar, what the report actually said is now becoming less familiar and often forgotten or misunderstood. On the occasion of its thirty year 'anniversary' it may be instructive to look in some detail at the actual wording of the report again (and this extensive summary includes therefore substantial extracts.)

The report itself is brief - a mere 34 pages - but it is supported by twenty appendices totalling a further 90 pages. Some of these appendices have had an impact as great as the report itself, in particular the evaluation study by John Carroll in Appendix 10.

The first point to note is that the report is entitled: **Languages and machines: computers in translation and linguistics**. It was supposedly concerned, therefore, not just with MT but with the broader field of computational linguistics. In practice, most funded NLP research at the time was devoted to full-scale MT.

The background to the committee is outlined in the Preface: "The Department of Defense, the National Science Foundation, and the Central Intelligence Agency have supported projects in the automatic processing of foreign languages for about a decade; these have been primarily projects in mechanical translation. In order to provide for a coordinated federal program of research and development in this area, these three agencies established the Joint Automatic Language Processing Group (JALPG)."

It was the JALPG which set up ALPAC in April 1964 under the chairmanship of John R.Pierce (at the time, of Bell Telephone Laboratories). Other members of the committee were John B.Carroll (Harvard University), Eric P.Hamp (University of Chicago), David G.Hays (RAND Corporation), Charles F.Hockett (Cornell University, but only briefly until December 1964), Anthony G.Oettinger (Harvard University), and Alan Perlis (Carnegie Institute of Technology). Hays and Oettinger had been MT researchers, although no longer active when ALPAC was meeting (having become disillusioned with progress in recent years); Perlis was a researcher in Artificial Intelligence; Hamp and Hockett were linguists; and Carroll was a psychologist. The committee did, however, hear evidence from active MT researchers such as Paul Garvin and Jules Mersel (Bunker-Ramo Corporation), Gilbert King (Itek Corporation and recently IBM), and Winfred P.Lehmann (University of Texas).

The committee agreed at the outset that support for research in this area "could be justified on one of two bases: (1) research in an intellectually challenging field that is broadly relevant to the mission of the supporting agency and (2) research and development with a clear promise of effecting early cost reductions, or substantially improving performance, or meeting an operational." ALPAC rejected (1), deciding that the motivation for MT research was the practical one of (2) alone. For this reason, ALPAC "studied the whole translation problem" and whether MT had a role in it.

The second point to note, therefore, is that the report concentrated exclusively on US government and military needs in the analysis and scanning of Russian-language documents. It was not concerned in any way with other potential uses or users of MT systems or with any other languages.

The first half of the report (pages 1 to 18) investigated the translation needs of US scientists and government officials and overall demand and supply of translations from Russian into English. ALPAC began by asking whether, with the overwhelming predominance of English as the language of scientific literature (76% of all articles in 1965), it "might be simpler and more economical for heavy users of Russian translations to learn to read the documents in the original language." Studies indicated that this could be achieved in 200 hours or less, and "an increasing fraction of American scientists and engineers have such a knowledge", and it noted that many of the available opportunities for instruction were underutilized (Appendix 2)

Next it looked at the supply of translations within government agencies (including those sponsoring MT research). They used a combination of contract and in-house translators. The committee was not able to determine the exact number of in-house translators, but it did establish that the average salary of translators was markedly lower than that of government scientists. Nevertheless, it found "a very low rate of turnover among government translators. Indeed, the facts are that the supply exceeds demand." At the time of the report, no post of government translator was vacant while there were over 500 translators registered in the Washington area (statistics in Appendix 8 of the report).

The committee was thus prompted to ask whether there was any shortage of translators. The Joint Publications Research Service, it found, had the capacity to double translation output immediately: out of 4000 translators under contract only 300 on average were being used each month. Likewise, the National Science Foundation's Publication Support Program was prepared to support the cover-to-cover translation of any journal which might be nominated for complete translation by any 'responsible' society. Appendix 6 recorded 30 journals being translated from Russian in this way during 1964. Since, some had very low circulations (Appendix 6), ALPAC questioned the justification for this virtually "individual service".

Indeed, ALPAC wondered whether there were not perhaps an excess of translation, on the argument that "translation of material for which there is no definite prospective reader is not only wasteful, but it clogs the channels of translation and information flow." What it found was that many Russian articles were being translated which did not warrant the effort: according to a 1962 evaluation, only some 20 to 30% of Russian articles in some fields would have been accepted for publication in American journals; furthermore the delays in publication of cover-to-cover translations reduced their value. The committee concluded that the main need was for "speed, quality, and economy in supplying such translations as are requested."

At this point, before considering MT as such, the report anticipated its conclusions with the bald statement (page 16): "There is no emergency in the field of translation. The problem is not to meet some nonexistent need through nonexistent machine translation. There are, however, several crucial problems of translation. These are quality, speed, and cost."

On quality, ALPAC stressed that it must be appropriate for the needs of requesters: "flawless and polished translation for a user-limited readership is wasteful of both time and money." But there were no reliable means of measuring quality, and for this reason ALPAC set up an evaluation experiment (reported in Appendix 10). This study by John B.Carroll evaluated both human and machine translations, and it has great influence on many MT evaluations in subsequent years. It was supplemented in Appendix 11 by a study from the Arthur D.Little, Inc. of MT errors, based on the system in use at the time at the Foreign Technology Division, i.e. the system developed by Gilbert King at IBM (virtually a word-for-word system.)

On speed, ALPAC saw much room for improvement: scientists were complaining of delays; the most rapid service (from JPRS) was 15 days for 50 pages; the NSF translation of journals ranged from 15 to 26 weeks; documents sent to outside contractors by the US Foreign Technology Division were taking a minimum of 65 days; and when processed by the FTD's MT system, they were taking 109 days (primarily caused by processes of postediting and production, detailed in Appendix 5).

On cost, ALPAC considered what government agencies were paying and this varied from $9 to $66 per 1000 words. In Appendix 9 calculations were made of cost per reader of the different forms of translation, including unedited output from the FTD system. These costs included the expenditure of time by readers. Assuming that the average reader took twice as long to read unedited MT documents as good quality human translation (based on the results of Carrol's evaluation in Appendix 10), it concluded that if documents are to be read by more than 20 persons traditional human translation was cheaper than MT. As for the costs of postedited MT, they would include posteditors proficient in Russian; ALPAC concluded that "one might as well hire a few more translators and have the translations done by humans... [or] take part of the money spent on MT and use it either (1) to raise salaries in order to hire bilingual analysts - or, (2) to use the money to teach the analysts Russian."

At this point, the report turned to "the present state of machine translation" (pages 19 to 24). It began with a definition: MT "presumably means going by algorithm from machine-readable source text to useful target text, without recourse to human translation or editing." And immediately concluded: "In this context, there has been no machine translation of general scientific text, and none is in immediate prospect."

Support for this contention, ALPAC asserted, came from "the fact that when, after 8 years of work, the Georgetown University MT project tried to produce useful output in 1962, they had to resort to postediting. The postedited translation took slightly longer to do and was more expensive than conventional human translation." Likewise, ALPAC regarded it as a failure that the MT facility at FTD "postedits the machine output when it produces translations."

However, the principal basis for its conclusion was the results of Carroll's evaluation exercise in Appendix 10. "Unedited machine output from scientific text is decipherable for the most part, but it is sometimes misleading and sometimes wrong... and it makes slow and painful reading." The report then printed (on pages 20 to 23) what it held to be "typical" samples of the "recent (since November 1964) output of four different MT systems." These were presumably those used in the evaluation exercise, but this was not stated explicitly. The four systems were from Bunker-Ramo

Corporation, from Computer Concepts, Inc., from the USAF Foreign Technology Division, and from EURATOM. The first would have beens the system developed by Paul Garvin after he left Georgetown in 1960. The Euratom system was the Georgetown University system installed in 1963 at Ispra, Italy. The FTD system was, as already mentioned, the one developed by Gilbert King at IBM, using his patented photoscopic store (a precursor of the laser disk). The Computer Concepts company had been set up by Peter Toma after he left the Georgetown project in 1962; the system illustrated was presumably AUTOTRAN, based in many respects on the SERNA version of the Georgetown system, and a precursor of SYSTRAN. Only the Euratom and FTD systems were fully operational at this time, the other two were still experimental prototypes - but this was not mentioned by ALPAC.

After reproducing the MT samples, the report continued: "The reader will find it instructive to compare the samples above with the results obtained on simple, selected, text 10 years earlier (the Georgetown IBM Experiment, January 7, 1954) in that the earlier samples are more readable than the later ones." Twelve sentences from the highly-restricted demonstration model (see MTNI#8, May 1994) are then listed, with the comment: "Early machine translations of simple or selected text... were as deceptively encouraging as "machine translations" of general scientific text have been uniformly discouraging."

There can be no doubt about the deficiencies and inadequacies of the translations illustrated but it was perhaps a little unfair of ALPAC to take the results of general-purpose MT systems (some still experimental) working from unprepared input (i.e. with no dictionary updating) and to compare them with the 'doctored' output from a small-scale 'demonstration' system.

ALPAC concluded this chapter by stating that is was very unlikely that "we will not suddenly or at least quickly attain machine translation", and it quoted Victor Yngve, head of the MT project at MIT that MT "serves no useful purpose without postediting, and that with postediting the over-all process is slow and probably uneconomical." However, the committee agreed that research should continue "in the name of science, but that the motive for doing so cannot sensibly be any foreseeable improvement in practical translation. Perhaps our attitude might be different if there were some pressing need for machine translation, but we find none."

At this point, ALPAC looked at what it considered the much better prospects of "machine-aided translation" (not, as it stressed, human-aided MT, but what are now referred to as translation tools). It had high praise for the production of text-related glossaries at the Federal Armed Forces Translation Agency in Mannheim (Germany) and for the terminological database at the European Coal and Steel Community, which included terms in sentence contexts - this was the precursor of EURODICAUTOM. (Further details were given in Appendices 12 and 13, pages 79-90). Its general conclusion was that these aids, primitive as they were, were much more economically effective in the support of translation than any MT systems.

The alternative it saw as postedited MT. However, it admitted that it could not "assess the difficulty and cost of postediting". Appendix 14 (p.91-101) reported on a study involving the translation of two excerpts from a Russian book on cybernetics, and the postediting of a an MT version of one of the excerpts. Interestingly, "eight translators found postediting to be more difficult than ordinary translation. Six found it to be about the same, and eight found it easier." Most translators "found postediting tedious and even frustrating", but many found "the output served as an aid... particularly with regard to technical terms." Despite the inconclusiveness of this study, ALPAC decided to emphasise the negative aspects in the body of its report, quoting at length the comments of one translator: "I found that I spent at least as much time in editing as if I had carried out the entire translation from the start. Even at that, I doubt if the edited translation reads as smoothly as one which I would have started from scratch. I drew the conclusion that the machine today translates from a foreign language to a form of broken English somewhat comparable to pidgin English. But it then remains for the reader to learn this patois in order to understand what the Russian actually wrote. Learning Russian would not be much more difficult." In hindsight, it is perhaps significant that ALPAC did not take cognizance of the fact that human translation, particularly of technical texts, is frequently revised before transmission to clients.

At the beginning of the next chapter "Automatic language processing and computational linguistics", ALPAC made one of its most often cited statements, namely that "over the past 10 years the government has spent, thorough various agencies, some $20 million on machine translation and closely related subjects." The statistics provided in Appendix 16 (p.107-112) reveal that by no means all this sum was spent on MT research in the United States. Firstly, the total includes $35,033 on sponsoring three conferences and $59,000 on ALPAC itself. Secondly, it includes $101,250 in support of research outside the United States (at the Cambridge Language Research Unit) and $1,362,200 in support of research under Zellig Harris at the University of Pennsylvania which even at the time was not considered to be directly related to MT. Thirdly, it lists global sums from the US Air Force, US Navy and US Army (totalling $11,906,600) with no details of the recipients of the grants. Evidence from elsewhere (details in Hutchins 1986:168) suggests that much of the funds were in support of developments in computer equipment rather than MT research (perhaps up to two thirds of the USAF grants). In brief, the funding of US agencies on US research in MT may well have been nearer $12-13 million than the frequently repeated $20 million stated by ALPAC. The sum was still large, of course, and ALPAC was right to emphasise the poor return for the investment.

The main theme of this chapter on "Automatic language processing and computational linguistics" was a consideration of the contribution of MT research to advances of NLP in general. Summarizing the more extensive findings in Appendices 18 and 19, it found that its effect on computer hardware had been insignificant, that it had contributed to advances in "computer software (programming techniques and systems)", but that "by far the most important outcome... has been its effect on linguistics." Here they highlighted insights into syntax and formal grammar, the bringing of "subtler theories into confrontation with richer bodies of data", and concluding that although "the revolution in linguistics has not been solely the result of attempts at machine translation and parsing... it is unlikely that the revolution would have been extensive or significant without these attempts." (This is a view which would certainly be disputed today.) However, despite this favourable influence, ALPAC did not conclude that MT research as such should continue to receive support; rather it felt that what was required was "basic developmental research in computer methods for handling language, as tools for the linguistic scientist to use as a help to discover and state his generalizations, and ... to state in detail the complex kinds of theories..., so that the theories can be checked in detail."

In the final chapter (p.32-33), ALPAC underlined once more that "we do not have useful machine translation [and] there is no immediate or predictable prospect of useful machine translation." It repeated the potential opportunities to improve translation quality, particularly in various machine aids: "Machine-aided translation may be an important avenue toward better, quicker, and cheaper translation." But ALPAC did not recommend basic research: "What machine-aided translation needs most is good engineering."

ALPAC's final recommendations (page 34) were, therefore, that research should be supported on:

"1. practical methods for evaluation of translations;

2. means for speeding up the human translation process;

3. evaluation of quality and cost of various sources of translations;

4. investigation of the utilization of translations, to guard against production of translations that are never read;

5. study of delays in the over-all translation process, and means for eliminating them, both in journals and in individual items;

6. evaluation of the relative speed and cost of various sorts of machine-aided translation;

7. adaptation of existing mechanized editing and production processes in translation;

8. the over-all translation process; and

9. production of adequate reference works for the translator, including the adaptation of glossaries that now exist primarily for automatic dictionary look-up in machine translation."

Aware that these recommendations failed to support not just MT but any kind of natural language processing, a statement was inserted in the final report addressed to the president of the

National Academy of Sciences from the chairman John R.Pierce in which he stressed the value of supporting "computational linguistics, as distinct from automatic language translation" Elaborating on recommendations in its chapter on NLP, the chairman believed that the National Science Foundation should provide funds for research on a reasonably large scale, "since small-scale experiments and work with miniature models of language have proved seriously deceptive in the past," - obviously alluding to MT experience -"and one can come to grips with real problems only above a certain scale of grammar size, dictionary size, and available corpus."

The ALPAC report was relatively brief; and its direct discussion of MT amounted to just one chapter (p.19-24) and four appendices (on evaluating translation (p.67-75), on errors in MT (p.76-78), on postediting MT compared with human translation (p.91-101), and on the level of government expenditure on MT (p.107-112)). The rest of the report was concerned with the demand for translation in general by US government agencies, the supply of translators, with computer aids for translators, and with the impact of MT on linguistics. However, it was in these few pages that ALPAC condemned MT to ten years of neglect in the United States (longer, as far as government financial support was concerned), and it left the general public and the scientific community (particularly researchers in linguistics and computer science) with the firm conviction that MT had been a failure or, at best, very unlikely to be a useful technology - a view which is still widely held.

In some respects, the impact of ALPAC can be exaggerated. MT research in the US did not come to a complete and sudden halt in 1966. Some projects continued, notably at Wayne State University under Harry Josselson until 1972 and at the University of Texas under Winfried Lehmann and Rolf Stachowitz until 1975 (later revived in 1978 with funding from Siemens). Furthermore, some MT projects supported by government money had ended before ALPAC reported: University of Washington (1962), University of Michigan (1962), University of Harvard (1964). In particular, the Georgetown University project, whose system was explicitly criticized by ALPAC, had received no funding after 1963. By this time it had installed operational MT systems at the Oak Ridge National Laboratory and at the Euratom laboratories in Italy.

Furthermore, in hindsight it can, of course, be agreed that ALPAC was quite right to be sceptical about MT: the quality was undoubtedly poor, and did not appear to justify the level of financial support it had been receiving. It was also correct to identify the need to develop machine aids for translators, and to emphasise the need for more basic research in computational linguistics. However, it can be faulted for concentrating too exclusively on the translation needs of US scientists and of US agencies and not recognizing the broader needs of commerce and industry in an already expanding global economy. In this way, ALPAC reinforced an Anglo-centric insularity in US research which damaged US research in multilingual NLP in general and allowed the lead in this field to be taken in Europe and Japan. It took two decades for the position to begin to be rectified in government circles, with the report for the Japan Technology Evaluation Center (JTEC 1992) and with ARPA support of current US research in this field.

References:

ALPAC (1966) *Languages and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences, National Research Council, 1966. (Publication 1416.) 124pp.

Hutchins, W.J. (1986) *Machine translation: past, present, future.* Chichester: Ellis Horwood.

JTEC (1992) *JTEC Panel report on machine translation in Japan.* Jaime Carbonell [et al.] Baltimore, MD: Japanese Technology Evaluation Center, Loyola College in Maryland, January 1992.

JEIDA (1989) *A Japanese view of machine translation in light of the considerations and recommendations reported by ALPAC, U.S.A.* Tokyo: JEIDA, 1989.

King, M. (1984) When is the next ALPAC report due? In: *10th International conference on computational linguistics...Proceedings of Coling84*, July 1984, Stanford University, Ca. (ACL, 1984), p. 352-353.

PUBLICATIONS: REVIEWED, ANNOUNCED, RECEIVED

Ovum' Report on Translation Technology Products: part II

Jackie Murgida

The first part of this review by Colin Brace appeared in MTNI#13: 2-3. This part of the review of Ovum's **Translation Technology Products** focuses on the quality, approach, and methodology of the evaluation sections of the report and should be read in conjunction with part one.

It may seem odd to describe the massive Ovum evaluation of translation technology products as an excellent beginning, but that is the impression given by a report that has a wealth of valuable information but leaves one wanting even more. The authors provide extensive information about commercial translation software, with an impressive amount of background detail which is absolutely necessary in determining whether products are appropriate for a given user environment and in choosing among the growing variety of available systems. Ovum has done a great service by emphasizing the complexity of linguistically sophisticated systems. They insist that users must evaluate their own specific situation in a holistic way and not just turn directly to the evaluations and follow the numerical ratings. They explicitly caution the reader: **not to compare products in different groups** (i.e., chapters), due to their different approaches and facilities; and **not to add up the rating scores** to get 'total scores' and thus determine a 'best product'. Choosing a translation technology product is much more complicated - a product that is suitable for one buyer's situation might be totally unsuitable for another.

Despite this excellent general approach, however, the reader is left with several questions, which, had they been answered, would have made the evaluations even more useful.

For instance, in the detailed evaluations of the "leading" products, the authors do not tell us how they decided which products to include in that category, nor do they explain the criteria for the numerical ratings assigned to the technical features of each system: usability, customisability, applicability to total translation process, language coverage, language processing technology, terminology building , and linguistic resources and utilities. The features are given a score on a ten-point scale, but there is no explanation of the criteria used to assign these scores or of what it would take for a product to get a score of ten (which none did). The reader must interpolate and infer this information from the accompanying discussion of the product and the company and by comparing the ratings received by all the other products in that category.

Similarly, while there is a discussion of "output quality" for each leading system, the authors state, "This section has not been included in the ratings but has been used to verify information provided by the vendor concerning linguistic depth, translation memory and terminology building." But they do not tell us why they limited the evaluation of output quality in this way.

For MT systems, the authors evaluated output quality using a set of 24-25 sentences which they put through the system to determine the number of meaningful sentences, the number of grammatically correct sentences, number of unknown terms and phrases, and the amount of time necessary to enter the unknown expressions. They then compared the output after terminology entry. All of this is excellent, invaluable information, but again, the reader does not know how the test sentences were developed -- from live texts or composed by the authors? isolated sentences or running text? what genre or text type? linguistically complex, average, or a range?

While we are told that the authors "interviewed" users, it is not clear to what extent the user's point of view was consulted. We read that the authors themselves did the testing and evaluation in

visits of two-to-three days at the company's development site and that Ovum has a "beta reviewer." There is no further information on the extent of the beta testing or the nature of the interviews with end users dealing with the systems in an operational setting.

More detailed information on the nature of the testing procedures and on the linguistic processing capabilities of the MT systems would have been highly useful for prospective users who need to post-edit raw output for publication-quality translations. For instance, the authors give information on the number of meaningful and meaningless sentences, but no explicit rating for accuracy. We must assume that in "meaningful" sentences the meaning in the output sentence matches the meaning in the input sentence. This is a reasonable enough assumption, but this type of information should be stated as part of the background on the evaluation procedures.

The information and explanations that are lacking from the report are greatly outweighed by the large body of useful data gathered in one volume, which is a great contribution to this field. The numerous caveats and guidelines on how to use the evaluations appropriately are especially important at a time when the people who make decisions on buying translation software sometimes are easily mesmerized by cutting-edge technology, regardless of how suited a product may be to the user's existing translation environment. The following advice is part of the report's "Messages to users":

> If you are involved in translation or need to provide multi-lingual information, translation technology products can help reduce your overall translation costs, improve quality and speed turnaround. If implemented properly, they will provide more accurate, consistent information and take much of the drudgery out of the translator's working day. However, if the wrong product is selected, it may cost a lot of effort and time to change course in mid-implementation. Choosing the right product should be taken every bit as seriously as selecting your computer hardware or word processing system.

This report helps you to determine the right product for your environment and how to implement it successfully. Ovum does not recommend that this report should replace an in-house evaluation of your current process and future requirements, but it will ensure that you consider all the important issues before deciding on a product. There are similarly cogent messages to suppliers.

The steep price of the report will make it difficult for individuals and smaller companies to obtain it. One can only hope that many potential users will have access to it somehow --perhaps through libraries -- and read it carefully, heeding the authors' excellent advice on using the information it contains.

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)

Melby on MT semantics

Review of:
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 14.) Amsterdam/Philadelphia: John Benjamins Publishing Company, 1995. 274 pp.

Klaus Schubert

Anamnesis
Unedited machine translation quality is poor for unrestricted text but quite good in neatly restricted domains. Alan Melby (in the philosophical chapter with Terry Warner) distinguishes superficial and fundamental ambiguity, i.e. homonymy vs. polysemy, and detects restricted-domain texts to lack fundamental ambiguity.

When modelled in semantic theory, superficial ambiguity suggests meaning to be describable as a set of atoms, as is often tacitly assumed in semantic representations for MT, whereas fundamental ambiguity leads to a view of vague, overlapping areas in a continuous space of meaning.

Present MT systems handle meaning, if at all, only inasmuch as it can be represented in distinct units. For a theoretical explanation Melby turns to the Generative school. In his account, Generative Grammar has always upheld the principles of studying sentences out of their textual connectedness and viewing syntax as autonomous from semantics and communicative use. The meaning of a sentence is received by bottom-up compositional interpretation. Thus, says Melby, Generative Grammar does not describe general language, but only the closely restricted domain of sentences which can be understood in isolation. MT based on this theory is successful where domain language meets this condition.

The Generative view of semantics is based on Chomsky's assumption of a supraindividual, objective, competence. According to Melby, this reduces the full infinity of expressible meaning to a predictable infinity which can be generated by algorithms. Melby suggests that this objectivist concept of a competence common to all speakers "totalizes" people, that is, "brings them under a single perspective by means of which the person who holds this view reduces them to replaceable role-takers in one vast system. When people are brought under this purview, the cognitive totalizing process is analogous to *political totalitarianism"* (p.119). Melby rejects this view as unethical and views Chomsky's philosophical stance as a reason for the failure of Generative Grammar (and MT systems based on it) in handling general-language text.

While Chomsky subsumes language acquisition under *individual* psychology, Melby views it as *social* interaction. Meaning is established and maintained through the interaction of many active language users, which is why meaning remains shifty and why creative metaphor (and thereby language change) is possible and ever-ongoing. Social interaction, however, is not computable - which is the ethical essence of Melby's argument.

Melby returns to MT, examining the implications of his philosophical plea. In language technology he encourages the development of restricted-domain and controlled-language MT and translation tools but recommends not to invest in FAHQT unless fundamental ambiguity is addressed. As for translation theory, Melby suggests dividing up the theory in two major branches along the same lines: one for general language and one for domain-specific controlled language. He then imagines MT to be successfully applied in a scenario where very strict control over the input language is possible. Out of this idea a third branch of translation theory emerges: text analysis with the task of establishing whether a text is machine-translatable.

Diagnosis

I find this book fascinating, stimulating and thought-provoking. Melby addresses the one issue decisive to any further progress in machine translation: the semantic barrier.

Melby's parallel between semantic objectivism and political totalitarianism may be felt offensive. But it is true that theories depicting meaning as something susceptible to predictable generating algorithms *reduce* the expressive power of the language of humans to that of abstract, objectivizable agents. While vagueness and dynamic change are constitutive and indispensable characteristics of language, these models describe something qualitatively much less expressive than human language.

Melby draws much on the difference, assumed fundamental, between domain and general language. However, his examples of translation problems are mostly occasional hits from journalistic and literary texts rather than systematic evidence from where the bulk translation work for industry and government is done. This could explain why Melby views the difference so absolute. In my view, it is a *gradual* difference. When a text is restricted to a single domain, most semantic decisions become superfluous because the number of options from which to choose in parsing, lexical transfer etc. is reduced to one. FAHQT is possible *to the extent* that semantics can be avoided. I therefore

should not split up translation theory: we need a single one which accounts for the *degree* to which the translation of various text types needs to resort to semantic (and extralinguistic) knowledge.

The book's merit lies in focussing our attention on the processing of meaning. The conclusions Melby draws from his analysis, however, would need more complete evidence.

Therapy

The classics teach us to view meaning as an infinite, unstructured continuum (e.g. Hjelmslev 1943: 48) to be mapped onto finite linguistic means (Humboldt 1836: 106). Semantic theories differ largely, but an accepted common notion is that of predication, which leads to models with structure-building and structure-filling elements. Therefore, a critical assessment of the syntactic basis of MT cannot be satisfactory without reference to the Dependency school with its straightforward correspondence between semantic predication relations and syntactic dependency lines, the more so since Dependency with its metataxis theory (Tesnière 1959: 283, Schubert 1987) has always focused on translation (unlike the Generative school, as Melby laments).

The infinity of meaning makes meaning difficult to represent other than by human language. Artificial representations do not achieve the expressive power needed to model human language. Therefore a series of interesting approaches attempt to keep meaning *implicit* (Schubert 1988). An evaluation of the semantic potential of MT would require a careful examination of the implicit analogical approaches, viz. the statistical, the example-based and the parallel-corpus techniques. For Melby's philosophical approach it must be an intriguing idea that by extracting knowledge from corpora one incorporates into MT an element of *observation* of the language community, whereby implicit meaning processing may exceed the limits of algorithmized, explicit and thereby reduced semantics.

Moreover, an investigation of semantics in translation needs to treat not only meaning in language as such, but to concentrate on the meaning-based decisions to be taken in the transition from one language to another, thus on lexical transfer and the semantic share in structural transfer.

Melby has written an inspiring book which refocuses our attention on the all-decisive issue in language technology. His drawing-up of the problem is called for and to the point. But his implications and his verdict on what and what not to fund cannot be taken at face value as long as they lack a thorough analysis of the implicit and analogical approaches to semantic processing and of Dependency syntax. Research goals ahead are above all decision mechanisms based on semantic and world knowledge interwoven with a syntax as isomorphic to semantics as possible. The book helps us clear our view for this goal.

References

Hjelmslev, Louis (1943): *Omkring sprogteoriens grundlæggelse.* New ed. 1966. s.l.: Akademisk Forlag

Humboldt, Wilhelm von (1836): *Über die Verschiedenheit des menschlichen Sprachbaues.* Berlin: Lambert Schneider / Königliche Akademie der Wissenschaften

Schubert, Klaus (1987): *Metataxis.* Dordrecht/Providence: Foris

Schubert, Klaus (1988): Implicitness as a guiding principle in machine translation. In: Dénes Vargha (ed.): *Coling Budapest.* Budapest: John von Neumann Society, 599-601

Tesnière, Lucien (1959): *Éléments de syntaxe structurale.* 2nd ed. 1982. Paris: Klincksieck

Klaus Schubert is professor of computational linguistics and technical translation at Fachhochschule Flensburg.

Analogical MT

Review of:
Jones, Daniel: ***Analogical natural language processing***. London: UCL Press, 1996. xii,155pp. ISBN: 1-85728-218-3

John Hutchins

This is the third volume in the series "Studies in Computational Linguistics" published in association with the Centre for Computational Linguistics at the University of Manchester Institute for Science and Technology, under the general editorship of Harold Somers. The first of the series was a reprint of *Linguistic and computational techniques in machine translation system design*, the 1983 report by Peter Whitelock and Kieran Kilby, which though now over ten year's old still retains value for the detailed description of the systems it covers. The second was *A course in generalized phrase structure grammar* by Paul Bennett, a textbook on a linguistic theory of interest to computational linguistics and hence to MT researchers.

The present volume is a monograph which aims "to challenge the currently predominant assumption in the field of natural language processing that the representation of language should be done within the rule-based paradigm alone." The author places his work, therefore, among the many other contemporary efforts in MT and computational linguistics to make use of statistical and analogical techniques based on large corpora of texts. Although intended more generally, the author describes research on translation to illustrate his arguments. Hence the book is likely to be of most interest to researchers in machine translation.

After an introductory chapter setting out aims and methodology, the author describes the 'background and relevant research' (p.10-38). In this, he describes the general principles of example-based machine translation (EBMT) and the basic features of the linguistic theory he will be using, the Functional Grammar (FG) developed by Simon Dik. The next chapter on 'Analogical translation' (p.39-68) is devoted to outlining a 'hybrid EBMT system' which employs FG-based representations to describe translation examples of both source and target languages. The fourth chapter on 'Stochastic and analogy-based language processing' (p.72-97) is a review of recent research on statistics-based MT and 'analogical modelling' techniques in NLP (particularly the connectionist model of Rumelhart and his colleagues); it is a survey of much value for those unfamiliar with these developments and their potential relevance to MT.

The fifth chapter entitled 'Experiments in analogical cloning' (p.98-141) presents the actual results of the author's research in four small-scale experiments. The first looked at the performance of analogical modelling (AM) to predict case roles of a small set of prepositions from business correspondence texts, and then in comparison with the performance of a connectionist model. The second experiment introduced some sentential analysis in order to measure lexical and phrasal similarities ('a simple cloning algorithm'). The third used AM to attempt to predict lexical categories and to assign case roles, demonstrating that rule-based morphological processing is unnecessary in an EBMT architecture. The fourth experiment was a refinement of the distance measurement using meaning definitions derived from the FG formalism, showing how information about synonymy may be used in example-based processing.

The final chapter (p.142-145) summarises how far the author has achieved his aim of demonstrating that analogical NLP can be a viable alternative to rule-based methods, and outlines the further work required. In sum, the volume is a clear presentation of the fundamentals of a potentially fruitful line of future MT research.

Language Engineering Directory: New Edition

The new *Language Engineering Directory* (1996 edition) is now available. The study from which it results is part-financed by DG-XIII of the Commission of the European Communities in Luxembourg.

The directory contains information on 700 organisations, 1,200 products and services in the language engineering field. Categories covered include: computer-assisted and machine translation, automatic indexing, document storage and retrieval, electronic dictionary products, handwriting recognisers, language and speech therapy products, multilingual word processing, optical character recognition, screen readers, speech recognition and synthesis, spelling, style and grammar checking, terminology management, thesauri, voice-assisted navigation and many others.

For further information, please contact: Language + Technology S.L., PO Box 2.572, 28080 Madrid, Spain. Tel: (34 1) 519 91 98, E-mail: 100575.2522@compuserve.com

---

Kluwer Annnounce Book Series:
Text, Speech and Language Technology

[Extract from publicity]
Due to the recent availability of large bodies of text and speech in electronic form, data-based research of all kinds has increased dramatically in areas such as computational linguistics and language engineering (especially corpus-based linguistics), speech, humanities computing, psycho-linguistics, and information retrieval. This series is intended to explore the methodologies and technologies that are emerging as a result of this work. In addition, while each of these disciplines has developed methodologies appropriate to its particular problem area, there is emerging a clearly defined set of technologies and methodologies common to all are as of research involving large quantities of electronic data. The series will be particularly concerned with methodologies and technologies with either actual or potantial applicability to other areas. The topics covered by the series include but are not limited to: encoding and representation of text and speech; lexical statistics and quantitative word studies; computational lexicography; morphological analysis and part-of-speech tagging; grammars and parsing technologies; automated content and thematic analysis; text databases and retrieval; document analysis, automatic indexing and abstracting; stylometry and computerized authorship discrimination; text generation; message understanding; text-to-speech and dictation systems; speech synthesis and speech recognition; phonological and prosodic analysis.

The series will be edited by Nancy Ide (Vassar College, USA) and Jean Véronis (CNRS and Université de Provence, France).

Forthcoming titles: *Text to Speech Synthesis* (T. Dutoit). --*Studies in Very Large Corpora* (ed. K. Church, P. Isabelle and D. Yarowsky). -- *Recent Advances in Parsing Technologies* (ed. H. Bunt and M. Tomita). -- *Prosody: Theory and Experiment* (ed. M. Horne). -- *Corpus-based Methods in Language and Speech* (ed. G. Bloothoft, K. Church and S. Young)

For more information contact: Polly Margules, Kluwer Academic Publishers, 101 Philip Drive, Norwell, MA 02061 (Tel: 617-871-6600 ext. 313; Fax: 617-871-6528; E-mail: pmargules@wkap.com)

Free Speech Journal

The *Free Speech Journal* is an electronic journal dedicated to fast publication of research advances and new ideas in areas of spoken language technology. *Free Speech Journal* offers peer review; feedback within two months of submitted manuscripts; immediate publication of articles accepted for publication; and all of the advantages of interactive electronic publishing. Our editorial board and reviewers' pool, with over two hundred experts in areas of human language technology, have each committed to fast review of manuscripts submitted for publication. In addition to publication on the Web, Free Speech Journal will be published annually as a book.

Free Speech Journal is now open for business. We welcome submission of manuscripts reporting original research and new ideas in the topic areas covered by the Journal. Information about the Journal and instructions for submitting manuscripts may be found at http://www.cse.ogi.edu/CSLU/fsj/home.html .

Integration of Natural Language and Vision Processing.
Edited by Paul McKevitt. Boston: Kluwer, 1995-96.

Although there has been much progress in developing theories, models and systems in the areas of Natural Language Processing (NLP) and Vision Processing (VP) there has heretofore been little progress on integrating these two subareas of Artificial Intelligence (AI). These four volumes contain a set of edited papers addressing computational models and systems for the integration of NLP and VP.

Vol. I: Computational Models and Systems. 1995. ISBN-0-7923-3379-9. Price: $US 110.00

Vol. II: Intelligent Multimedia. 1995. ISBN-0-7923-3758-1. Price: $US 110.00

Vol. III: Theory and Grounding Representations. 1995. ISBN-0-7923-3944-4. Price: $US 110.00

Vol. IV: Recent Advances. 1996

Contact: Polly Margules, Kluwer Office, Boston (USA) E-mail: kluwer@world.std.com; Tel: +1 617 871 6600; Fax: +1 617 871 6528

MUC Workshops

The Proceedings of the MUC Workshops are distributed by Morgan Kaufmann Publishing Inc. of San Francisco. There was no MUC in 1994. MUC-4 took place in 1992, MUC-5 in 1993 and MUC-6 in 1995. The MUC-6 Proceedings are still not available (the conference was in November) but should be out very soon. Distributor in UK: Direct Distribution, 27 Church Road, Hove, East Sussex, BN3 2FA, England (Tel: +44 1273 748427; Fax: +44 1273 722180; Email: dirdist@erlbaum.co.uk.

Survey of the State of the Art of Human Language Technology

This book is now available at http://www.cse.ogi.edu/CSLU/HLTsurvey/ .

The survey consists of articles by 97 authors in the following chapters: Spoken Language Input; Written Language Input; Language Analysis and Understanding; Language Generation; Spoken Output Technologies; Discourse and Dialogue; Document Processing; Multilinguality; Multimodality; Transmission and Storage; Mathematical Methods; Language Resources; Evaluation

The survey was funded by the National Science Foundation and the European Commission, with additional support provided by the Center for Spoken Language Understanding at the Oregon Graduate Institute and the University of Pisa.

PUBLICATIONS RECEIVED

Journals

AAMT Journal *no.12, October 1995* Contents on English diskette: On the AAMT/IAMT Bulletins (Hirosato Nomura). -- Nomura Laboratory, Kyushu Institute of Technology. -- New products: Mr.EJ Dic Reader, Netscape/ej, The Translater-Mini for Macintosh [see page ?]; *no.13, January 1996* Contents on English diskette: An unfinished dream of machine translation (Hozumi Tanaka). -- Nagao Communication Engineering Laboratory, Kyoto University (Makoto Nagao). -- New products: PENSEE for Internet for Windows, Rupo JW-V700, Japanese-English Dictionary Referencer, Dr.SURF for Windows [see page ?]

ALPNET news *issue 1, 1996.*

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

Elsnews *vol.5 no.1 (February 1996)*. Contents include: Transnational English database (Lori Lamel). -- Fourth European Summer School on Language and Speech Communication, Budapest July 1996 [announcement]. -- PARS and PARS/U: translation tools for Russian and Ukrainian (Michael S.Blekhman).

INL Infoterm Newsletter *77/78 (December 1995)*

Language Industry Monitor *no.26 (March-April 1995)*. Special issue on Speech recognition.

Language International *vol.8 no.1 (Februray 1996)*. Contents include: Automatic translation by micro-computer: evaluation of Power Translator (Laurence Nazé). -- Accent adds bidirectional processing. -- Software localisation in Ireland, SLIG '95 (Reinhard Schäler). -- MT Summit V, Luxembourg July 10-13 1995 (Robert Clark). -- European Commission: a translation service with a view (Thierry van Steenberghe). -- MT on-line from CompuServe [includes interview of Mary Flanagan by Geoffrey Kingscott]; *vol.8 no.2 (April 1996)*. Contents include: METAL at Boehringer Ingelheim (Geoffrey Kingscott). -- Logos Corporation.

LISA Forum Newsletter *vol.5 no.1 (February 1996)*. Contents incllude: LISA, the vision for localisation IT - what is it? (Judith Jones). -- Translation and the Internet (Robin Bonthrone). -- Transparent language processing: a solution for internationalization of Internet services (Borka Jerman-Blazic et al.) -- The multilingual World Wide Web (Gavin Nicol). --Multilingual API for internationalization of the network services (Borka Jerman-Blazic and Andrej Gogala).

Literary and Linguistic Computing *vol.11 no.1 (April 1996)*

Machine Translation *vol.10 nos.4, 1995*. Contents: pp. 269-291: A model of a bi-directional transfer mechanism using rule combinations (Hideo Watanabe). -- pp.293-319: Parsing Turkish using the Lexical Functional Grammar formalism (Zelal Güngördü and Kemal Oflazer). -- pp.321-341: Book reviews

Natural Language Engineering *vol.1 part 3 (September 1995)*. Contents: Anaphora and ellipsis in artificial languages (Stephen G.Pulman). -- Russian morphology: an engineering approach (Andrei Mikheev and Liubov Liubushkina). -- Syntactic analysis of Hebrew sentences (Shuly Wintner and Uzzi Ornan). -- A hierarchical Dirichlet language model (David J.C.MacKay and Linda C.Bauman Peto).

Terminology *vol.2 no.2 (1995)*

Terminology Standardization and Harmonization: newsletter of ISO/TC 37, *vol.7 no.3/4 (December 1995)*

The Translator *vol.2 no.1 (1996)*

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 on page ?]

W.John Hutchins and Harold L.Somers: **Introducción a la traducción automática**. Trad. dirigida por Joseba K.Abaitua Odriozola. (Lingüística y Conocimiento 21). Madrid: Visor, 1995. 480pp. ISBN: 84-7774-871-3.

Daniel Jones: **Analogical natural language processing**. (Studies in Computational Linguistics). London: UCL Press, 1996. xii,155pp. ISBN: 1-85728-218-3. [Review on page ?]

Conference proceedings

CLAW 96: Proceedings of the First International Workshop on Controlled Language Applications, 26-27 March 1996, Centre for Computational Linguistics, Katholieke Universiteit Leuven, Belgium. iv,242 pp. [See page ?]

Reports

Linguistics Research Center, The University of Texas: Quarterly report, January-March 1996. 19 pp. Contains review by W.P.Lehmann of first three volumes in series *Studies in Computational Linguistics* (ed. Harold Somers), UCL Press.

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 [Reviews MTNI#13: 2-3, and page ? of this issue]

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.