MT News International

Newsletter of the International Association for Machine Translation

ISSN 0965-5476 Issue no.12, October 1995

IN THIS ISSUE:

[Note: this electronic version contains some long items omitted from the printed version]

Editor-in-Chief:
John Hutchins, The Library, University of East Anglia, Norwich NR4 7TJ, United Kingdom
Fax: +44 (1603) 259490; Email: J.Hutchins@uea.ac.uk; or: 100113.1257@compuserve.com
Regional editors:
AMTA: Joseph E.Pentheroudakis, Microsoft Corporation, Redmond, WA 98052, USA
Fax: +1 (206) 936-7329; Email: josephp@microsoft.com
EAMT: Jörg Schütz, Institute for Applied Information Sciences (IAI), Martin-Luther-Str.14, D-66111 Saarbrücken, Germany. Fax: +49 681 397482; Email: joerg@iai.uni-sb.de
AAMT: Professor Hirosato Nomura, Kyushu Institute of Technology, Iizuka, 820 Japan.
Fax: +81 (948) 29-7601; Email: nomura@dumbo.ai.kyutech.ac.jp
Advertising Coordinator:
Bill Fry, Association for Machine Translation in the Americas, 2101 Crystal Plaza Arcade, Suite 390, Arlington, VA 22202-4616, USA. Tel: +1 (703) 998-5708; Fax: +1 (703) 998-5709.
Published in the United States by Jane Zorrilla

CONFERENCE REPORTS

MT Summit V held in Luxembourg
10-13 July 1995

John Hutchins

The fifth of the biennial MT Summits was held in the Hemicycle of the European Parliament from 10th to 13th July. Organised by the SEMA Group on behalf of the European Association for Machine Translation, the conference was supported financially by the European Commission.

The conference was preceded by tutorial sessions held primarily for participants not familiar with MT and recent developments. These sessions were given by Margaret King (an introduction to MT), by Colin Brace (MT and its users), by Muriel Vasconcellos and Karin Spalink (MT in practice), and by Harold Somers (current research developments).

The conference proper was opened by two speakers from the Commission (Vicente Parajon Collada of DG XIII, and Eduard Brackeniers of the Translation Service), who both stressed the vital significance of MT and language engineering in general for the future success of the European Union. In particular, Dr Brackeniers believed that standardisation and automation were the only sure ways of saving the EU from its own bureaucratic paper mountains. The formal opening of the conference was performed by Margaret King, president of IAMT, and by Yorick Wilks, the chair of the programme committee.

At this MT Summit, presentations were restricted in the main to fifteen minutes, which encouraged succinctness at the occasional expense of otherwise useful detail and, more significantly, of opportunities for comments and questions, but with the advantage that participants from many backgrounds were exposed to the wide range of activities and issues which characterise MT at the present time.

An initial session looked back at some aspects of the past of MT: Bente Maegaard drew lessons from Eurotra and the successful implementation of PaTrans, John Hutchins reflected on some lessons from MT history, Christian Boitet examined some reasons for successes and failures in MT, and Dale Bostad recounted the experience of USAF and the NAIC in over 25 years as MT users.

As to be expected, there were number of presentations of commercial and operational MT systems with emphases on recent developments: Metal (Thomas Schneider), Logos (Scott Bennett), ATLAS (Masaru Fuji), ETAP (Igor Boguslavsky), Duet (Yoji Fukumochi), JICST (Tatsuo Ashizaki), and the recently launched Personal Translator PT (Hubert Lehmann, IBM). In addition, all the major manufacturers were to be found in the exhibition area: AppTek, Hypertrans (Dagoservice), Globalink, ISS (Singapore), Eurolang, Logos, ALEP, Alethtrad (GSI-Erli), Systran, Transit (STAR), METAL (GMS and Sietec), and Trados.

There were even more presentations which focused on the experiences of MT users. A general overview was provided by Muriel Vasconcellos and Colin Brace in a follow-up of the wide-ranging survey given at the previous MT Summit [the complete article is included in this issue of MTNI on page ??]. Adriane Rinsche gave a summary of the main findings of the soon-to-be-published Ovum reports on the market for MT technologies [see also page ? in this issue]. Each stressed the significant trend towards PC-based software, towards use by non-translators and to the growing influence of networking for both large and small users.

One whole session was devoted to the use of MT and translation tools within the Commission: Roberto Cencioni spoke about the general area of language engineering, Dorothy Senez described the recent rapid increase in the use of Systran - primarily for information assimilation - and Jean-Marie Leick outlined the development of EURAMIS (European Advanced Multilingual Information System) which will combine translation memories, terminology databases and machine translation facilities into an integrated work station available for EC staff in late 1996.

Other speakers recounting their experience as MT users included Susumu Donomae of Nagase (Japan) on practical experience with MT pre- and post-editing and translators' reactions; Doris Marty-Albisser on the installation of MT at the Union Bank of Switzerland; Bernard Masion on the cost-effective use of the METAL system at the Siemens Nixdorf; Daniel Grasmick described experience of the SAP software company with Logos and later with METAL, and Peter Wheeler spoke about the use of Logos at his MT bureau service. But the newest area of application was that of the Intergraph system for translating messages on the CompuServe network; Mary Flanagan emphasised the challenge for MT to deal with transient text, often ungrammatical, badly punctuated and highly variable in domain, and when the results are wanted immediately. CompuServe is not alone now: Globalink has recently introduced an Internet service, and of course network access has been available for some time to Systran and ATLAS.

What these developments illustrate, above all, is the need for reliable evaluations and publicly accepted benchmarks. One session was devoted to evaluation: Hitoshi Ishara described the JEIDA test suites for English-Japanese systems; Elliott Macklovitch spoke about the development at CITI of the TransCheck tool for translators to check for terminology consistencies, omitted sentences, false cognates, etc.; John White described the methods used to evaluate the DARPA-sponsored systems Pangloss, Candide and Lingstat in comparison with other operational systems; and Sylvie Regnier-Prost spoke on the procedure adopted at Aerospatiale to evaluate potential systems for their own needs.

Another major focus was the development of language resources for MT and related fields. Brian Oakley sketched the background to the establishment of the European Language Resources Association [see elsewhere in this issue], including the experience with Eurotra. Kazunori Muraki described the large EDR project in Japan; Christian Galinski spoke on terminology resources; Gudrun Magnusdottir argued that the data resources built by MT projects may be more useful to language technology in general than the MT systems themselves; and Zaharin Yussoff reported on the establishment in Malaya of a centre for language data resources (including a translation service).

The future of MT research was the topic of papers from Makato Nagao and Jun-ichi Tsujii, and Sergei Nirenburg told participants how to survive in the competitive world of research contracts. One session was devoted to progress in the area of MT research with the highest profile at the present time, namely speech translation. Wolfgang Wahlster reviewed the wide range of activities in the Verbmobil project (including five rival groups working on speech recognisers); Hitohi Iida, in a general survey of MT for spoken language, brought us up to date with progress on the ATR research; and Alex Waibel reported on the achievements so far on JANUS and the C-STAR consortium. All projects are ambitious efforts to integrate speech technology and MT, but all are restricted to the relatively narrow domains.

The wider integration of MT and other documentation process were covered in the final papers of the conference. The problems of multilingual searching of databases and of multilingual information retrieval were the topic for Gregor Thurmair, who outlined a project which combined MT, text extraction, fuzzy matching, translation memories, and report generation. Tsuyoshi Morimoto presented a future of multimodal multimedia telecommunications, which would necessarily also be multilingual and require MT facilities. Both Key-Sun Choi and Harald Zimmermann speculated on the future integration and standardisation of MT in office automation environments.

The conference was brought to a close by Jan Roukens of the European Commission who

reiterated the central role of language technology in the multilingual European information society and stressed the importance of promoting the credibility, acceptability and economic viability of MT among senior policy makers. Finally, Muriel Vasconcellos as the next president of IAMT invited everyone to the next MT Summit which will be hosted in two years' time by the Association of Machine Translation in the Americas. [For details of the planning see Conference Announcements.]

After the close, participants were invited to a demonstration by the Translation Service of the Commission in the Jean Monnet building. The presentation highlighted the role of automation in the service and future plans for the EURAMIS project, terminology analysis, full text searching of translation corpora, and an integrated workbench for in-house translators.

A selection of papers from the proceedings of the conference was compiled by the SEMA Group Belgium S.A. (96 rue de Stalle, B-1180 Brussels, Belgium; fax +32-2-333-5522).

TMI-95 in Leuven, Belgium
5-7 July 1995

John Hutchins

The sixth of the international conferences on Theoretical and Methodological Issues in Machine Translation (TMI-95) was held in Leuven (Belgium) in the week preceding the MT Summit conference, from the 5th to the 7th July. It was organised by the Centre for Computational Linguistics of the University of Leuven under the chairmanship of Geert Adriaens, with Frank van Eynde as head of the program committee. This TMI was the first to be held in Europe: the first four were North American events, and the fifth took place in Kyoto in conjunction with the MT Summit in 1993. As in previous years, whereas the MT Summit is an occasion for the MT community to display itself to a wide public, the TMI offers researchers the opportunity to present the latest developments in the field.

The organisers devised a programme of alternating longer and shorter papers, with two invited talks. The first by Manny Rayner (SRI) was a review of research on speech translation and its major challenges, concentrating on experience from five projects: SLT (a joint project of SRI, Telia Research, SICS, and ISSCO), Verbmobil (at a number of German universities), JANUS (Carnegie Mellon University), ASURA (the ATR system), and BLAST (at AT&T). Manny Rayner stressed that the interface between the speech technologies and MT has proved easier than expected. Nevertheless there remained familiar problems of noisy input, lack of pre- and post-editing, lack of visual feedback, and low correction rates from parsers. Although each of the systems adopts a different set of methods, it is found that every system is good at the same sentences and that every one is bad at the same problems. The most hopeful approaches include the pipelining of components, transfer at deep syntactic and shallow semantic level, hybrid (rule- and statistics-based) language processing, fast processing, high-quality polyphone synthesis; he identified the next area for research focus as the modelling of discourse phenomena.

Speech translation was the topic of a number of papers. Four concerned aspects of Verbmobil project. Dan Flickinger spoke about an implementation of minimal recursion semantics within HPSG; Birte Schmitz on the analysis of dialogue acts; Joachim Quantz on the domain model for Verbmobil; Jan Amtrup on a method of incremental transfer appropriate for speech translation.

Two papers were reports of the JANUS project, one from Alex Waibel evaluating the effectiveness of a variety of different parsing techniques (rule-based, statistical, discourse plans) in

disambiguation; and the other from L.J.Mayfield on the parsing of spontaneous speech as sequences of concepts rather than as sentences.

The ATR project was represented by a paper given by Eiichiro Sumita on the use of parallel machines for obtaining best-match partial translations in their example-based approach. J.M.Vilar from the University of Valencia outlined an approach to continuous speech translation in a limited domain based on a finite-state transducer.

The application of statistical techniques is now beginning to dominate many MT conferences, and TMI was no exception. Yasuhiro Akiba described the NTT experiments in automatic acquisition of English verb selection rules from a combination of real and hand-made example sentences; the results were evaluated in the ALT Japanese-English system. Ralf Brown and Robert Frederking described the construction and application of a statistical language model to select between alternate translations produced by the three MT engines (transfer-based, knowledge-based, and example-based) in the Pangloss project. Pascale Fung introduced a 'coerced Markov model' to relate lexical sequences of source languages and tagged sequences of target languages as part of a project towards a statistical transfer-based MT system. Kuang-hua Chen described the hybrid approach being explored at the National Taiwan University in an English-Chinese system. Naohiko Uramoto spoke on the work at IBM Japan for deriving frequency preferences for translating noun compounds in an example-based approach. The most extreme statistical approach was a proposal from Dekai Wu involving the construction of a probabilistic translation lexicon from an aligned bilingual corpus without any language-specific grammatical information.

More traditional in subject matter were the presentations by Oliver Streiter on problems of derivational morphology and syntax (part of the CAT2 project), and by Claudia Gdaniec on structural transfer in the Logos English-German system. Frank van Eynde described an HPSG approach to the problems of translating temporal expressions. Sergei Nirenburg argued the merits of interlingua approaches against the content-less formalisms of recent years. The lexicalist 'shake-and-bake' conception was the topic of two speakers: Davide Turcato spoke on morphological issues, and Arturo Trujillo on bilingual lexical transfer in the model. Three papers looked at problems of anaphora and reference: Ruslan Mitkov in the context of the CAT2 model; Francis Bond on the approach to noun phrase reference in ALT/JE, and Hiromi Nakaiwa on Japanese zero pronouns for the same system.

The second invited talk was given by Edward Johnson (University of Cambridge) who recounted his experience in the development of controlled languages for the police, maritime communication, illustrated by numerous humorous examples of misunderstandings in real life.

A number of other papers concerned domain restriction, sublanguages and controlled language. Geert Adriaens reported on an evaluation of the SECC grammar checker and corrector. Guy Deville described the ANTHEM sublanguage system for a multilingual healthcare information system. Teruko Mitamura spoke about the design of the controlled language in the KANT system under development at Carnegie Mellon University for the Caterpillar Corporation. Setsuo Yamada described a method of adapting the ALT/JE system to different domains. Domain adaptation was the topic also of the contribution of Key-Yih Su, which put forward the combination of a variety of rule-based and statistical techniques for parameterizing MT systems for different clients and domains by the automatic acquisition of transfer and translation knowledge from bilingual corpora.

At the conference dinner, Sergei Nirenburg entertained participants with recollections of the first TMI meetings, and invited everyone to come to the next TMI on its return to North America

when it is to be held at New Mexico State University in 1997.

The proceedings are available from the Centre for Computational Linguistics, Katholieke Universiteit Leuven, Maria-Theresastraat 21, Leuven, Belgium (Tel: +32-16-325088; Fax: +32-16-325098; Email: tmi95@ccl.kuleuven.ac.be; http://www.ccl.kuleuven.ac.be)

Third Workshop on Very Large Corpora
30 June 1995

Ken Church

The Third Workshop on Very Large Corpora was held at MIT on June 30, 1995, just after ACL-95. Every year we keep saying that corpus research is a hot topic, and every year the crowds just keep growing. But everyone (including the caterer) were surprised by the large turnout this year, almost 200 people, more than twice as many as last year.

The MIT location was excellent and is at least partly responsible for the large turnout. It was a bit ironic, though, to hold a conference on empiricism at MIT given MIT's long-time association with alternative perspectives.

Without question, the high point of the day was the invited talk by Henry Kucera and Nelson Francis. The Brown Corpus is still cited in practically every paper, even though it is nearly 35 years old. Many of us are beginning to appreciate that the Brown Corpus has had a longer-lasting influence on the field than many other topics from the period, many of which received far more attention and financial support (especially at MIT).

The reaction to Kucera and Francis' was extraordinary. Many of us looked back at our own research and realized that the Brown Corpus had not only changed the field, but it had changed us as individuals. For me personally, I suddenly realized that I had out-grown the ways of my MIT teachers and have adopted a new/old/different direction that that I could identify with. Many others in the audience must have shared similar feelings. The talk ended with a standing ovation (and endless requests for autographs), something you don't often see in a technical conference.

The proceedings of the Third Workshop on Very Large Corpora are now available for sale by mail order. The cost is $25, which includes first class postage. Checks and Visa/Mastercard are accepted as payment. To order proceedings, please send a completed order form to: Priscilla Rasmussen, ACL Office Manager, P.O. Box 6090, Somerset, NJ 08875. Email orders may be sent to acl@bellcore.com

Contents:

ZHOU/DAPKUS: Automatic Suggestion of Significant Terms for a Predefined Topic

RILOFF/SHOEN: Automatically Acquiring Conceptual Patterns without an Annotated Corpus

CHEN/LEE: Development of a Partially Bracketed Corpus with Part-of-Speech Information Only

FUNG: Compiling Bilingual Lexicon Entries From a Non-Parallel English-Chinese Corpus

MELAMED: Automatic Evaluation and Uniform Filter Cascades for Inducing N-Best Translation Lexicons

---

Recent Advances in Natural Language Processing (RANLP'95)
14-16 September 1995

Nicolas Nicolov
(University of Edinburgh)

The conference took place in the scenic area near the Batak Lake in Southern Bulgaria. The aim of the conference was to allow researchers to present new results in NLP based on modern theories and methodologies. Alternative techniques to the mainstream symbolic NLP like analogy-based, statistical and connectionist approaches were covered too. Similar in spirit were the lectures of the summer school on "Contemporary Topics in Computational Linguistics" which was held immediately before the conference (lecturers: J.Tsujii, J.Hutchins, A.Joshi, Z. Yusoff, Y.Matsumoto, K.Choi, R.Mitkov, A.Ramsay, M.Kudlek and W. von Hahn). In fact many people attended both events.

The three days of the conference were full of presentations – the proceedings include 46 papers (32 papers and 14 project notes) out of well over a hundred submissions. The event attracted authors from 18 countries (18.7% from UK, 14.5% from Japan, 12.5% from Germany, 12.5% from France, etc.). Each day started with an invited talk: A.Joshi discussed "Linguistic, computational and statistical implications of lexicalised gramars"; J.Tsujii talked about "MT: productivity and conventionality of language"; C.Boitet addressed "Ambiguities and Ambiguity labeling: towards ambiguity databases". The majority of the papers were on applying corpora-based and statistical methods to subtasks of MT or building knowledge sources for MT using data-oriented techniques. MT has again proved to be a fruitful testbed for research in computational linguistics. The wide availability of linguistic resources in electronic form has made it possible for researchers to address problems about which in the past it was only possible to make guesses and has allowed for current systems to have realistic performance (optimised algorithms). There were also papers on generation, parsing and morphology proper. Less conventional topics like the treatment of irony were also addressed. A trend that was observed is the strong element of multilinguality in current research. The best presentation was by Hideki Kozima on "Contex-sensitive measurement of word distance by adaptive scaling of a semantic grid". The papers from the conference will appear as a book to be printed by an international publishing house.

In case you believe everything has been planned by the organisers here is something that will convince you otherwise. Day two ended by a mock presentation by C.Samuelsson on "Missing Words have Meaning in Context" (jointly concocted by) R.Basili, M.Owens, C.Samuelsson and K.Simatan. All had a good laugh when we were most tired. The presentation (with proper slides, references, etc, even questions) was a kind of a humourous follow-up after Owens' presentation

which closed day one and which provoked an interesting discussion. Apparently the above authors have been discussing the topic over the drinks during the reception and the initial plan was written on a napkin.

In addition the conference included presentations of NLP software -- Mark Brissenden demonstrated Glosa International's environment for developing an English lexicon.

The conference included a day trip to old Plovdiv (second biggest city in Bulgaria). RANLP'95 was held in the former rest house of the Ministry of Foreign Affairs. We believe we have created a good working atmosphere which has given the participants maximum opportunity for informal discussions in the beautiful Rhodope mountains near the Batak lake. The idea to have a shuttle bus service from Sofia airport directly to the conference site proved good and saved time for the participants many of whom had a long journey to Bulgaria.

The conference is expected to be held biennially -- stay tuned and be on the alert for the next announcement. Information about RANLP'95 including titles of the papers, short abstracts, addresses of participants, etc. is available on the web at: http://www.dai.ed.ac.uk/misc/NLP-Conf.html

---

Second ALEP User Group Workshop
18-19 July 1995

Jörg Schütz

The second AUG workshop took place short after the delivery of the ALEP 2.3 release by Cray Systems. The beta-tests of this system where promising with respect to the overall behaviour of the system (stability, time and user-friendliness), which was also reported by some of the participants of the workshop. Like the first AUG workshop, this workshop was also a very successful event in terms of lively discussions and fruitful exchanges between the participants.

The content of the workshop can be divided into three main categories:

ALEP development related issues.
Grammar Engineering with ALEP and ALEP-based applications.
Language Engineering (LE) in general.

In the following we give a brief overview of the talks of the second AUG workshop. All talks will be published in volume 10 of the series ‘Studies in Machine Translation and Natural Language Processing’ of the European Commission.

The workshop started with an overview of the ongoing work of the ALEP Support and Maintenance project (LRE-62101) by Jörg Schütz (IAI, Saarbrücken) and Neil Simpkins (Cray Systems, Luxembourg). The main future subject of this project will be the port of ALEP to Quintus and Sicstus. The port to both platforms is feasible because Sicstus 3.x is much closer to Quintus than previous Sicstus releases. A beta-version is foreseen for mid-September and the actual delivery of the Quintus/Sictus ALEP system will start in the beginning of October. In addition, it is also planned to release a “streamlined” PC-version of ALEP (cf. the ALEP Shop SIG below).

Paul Schmidt (IAI, Saarbrücken) started a series of LS-GRAM related talks. He gave a brief overview of the progress of the project, particularly he reported about his first experience with the new ALEP release and compared them with the results he obtained with previous ALEP releases (cf.~summary of first AUG workshop, AUG Newsletter No.~2). His results are very convincing, and it seems that the ALEP system has reached a stability where it is mature enough for real-life

Language Engineering.

Brad Music (CST, Copenhagen) reported about their work in text pre-processing in the PaTrans system and how similar approaches could be integrated into the ALEP texthandling subsystem.

Andrew Bredenkamp (University of Essex) gave an overview of their English HPSG implementation in the LS-GRAM project. His focus was in particular how specific HPSG constructs can be expressed or simulated in the ALEP formalism.

Montserrat Marimon (Fundacio Bosch-Gimpera, Barcelona) talked about the implementation of a HPSG inspired grammar for Spanish in the ALEP formalism. Like Andrew, she also focused on specific implementation issues.

The afternoon sessions were dedicated to the integration potential of ALEP. Flora Ramirez-Bustamente (Universidad Carlos III, Madrid) gave an overview of the GramCheck project, which intends to deliver a grammar and style checker for Spainish and Greek based on the ALEP system. She briefly introduced their approach to grammar and style checking, and outlined how this approach was integrated in the Spainish LS-GRAM grammar.

Paloma Martinez (Universidad Carlos III, Madrid) reported about the actual software integration of the GramCheck modules in the ALEP environment.

The first workshop day ended with a presentation by Marius Groenendijk (Cray Systems, Luxembourg) about the general integration potential provided by the ALEP system.

The second day started with a talk by Thierry Declerck (IMS, Stuttgart) about the treatment of information across sentence boundaries. His approach is based on the so-called Dynamic Predicate Logic (DPL), and his current realization is facilitated by the texthandling subsystem and the analysis subsystem of ALEP. This work is also conducted within the LS-GRAM project.

Axel Theofilidis (IAI, Saarbrücken) presented his approach of dealing with frame variation in a HSPG inspired grammar for German through the use of a co-representation of head information. This work is also related to the LS-GRAM project.

Then the presentations switched to more general issues in Language Engineering. Cornelius Koster (University of Nijmegen) gave an introduction to the AGFL (Affix Grammars over Finite Lattices) system, in particular he advocated for a coupling of AGFL and ALEP in order to have better performance results for parsing. He also gave a demonstration of his system, where it became obvious that the AGFL system lacks appropriate linguistic sophistication.

Jörg Schütz (IAI, Saarbrücken) presented an overview of the evolution of the term Language Engineering, and presented a first approach towards a common methodology for LE.

Finally, Key-Sun Choi (KAIST, Korea) gave an overview on how his institution views LE, and how the recently initiated Korean LE programme intends to reach the goal of a Korean Language Industry. Currently, KAIST is the only institution with an industrial basis, which expressed their interest in ALEP from an industrial point of view.

ALEP Special Interest Groups

After the presentations, the ALEP-2.3 system was demonstrated to those participants who are newcomers in the ALEP community. The other participants split into two so-called ALEP Special Interest Groups (ASIGs). A summary of the results of the two ASIGs, Texthandling and ALEP Shop, is given below.

SIG ALEP Shop

The main focus of the ALEP Shop SIG was the development of an ALEP PC version, in

particular the look and feel of its user interface and the provided functionalities. In principle, this interface should be as simple as possible because the PC version is not meant as a development platform. ALEP lingware development still shall be conducted under Unix. The main purpose of the PC version version shall be for demonstration and education, and for running applications. However, the latter might be difficult because most applications are based on the ALEP User Language, and the port of the AUL is not yet intended for the PC version.

As a potential interface we discussed an implementation under MS Windows or just a simple command interpreter running in a DOS-shell (comparable to the ALEP0 command interpreter); as a minimal set of commands we identified: load, analyse, refine, translate and synthesis.

A Xmfed tool with a reduced functionality was seen as being an asset for the ALEP PC version.

Neil Simpkins of Cray Systems will prepare a short discussion paper on a possible architecture of the PC version, which will be distributed via the AUG mailing list.

An additional subject of the SIG discussion was about additional features of the ALEP Unix version. Here, multiple inheritance was the main issue. If MIH is supported then there should be also the possibility to view graphs with the Xmfed tool.

A last issue of this SIG was the future support, maintenance and development of the ALEP system. It was suggested to establish something similar to the GNU group to ensure these kinds of activities. Currently, this might be problematic (because of the BIM licenses) but with the Quintus/Sicstus version of the system it will be more feasible.

SIG Texthandling

The TH SIG discussed aspects of an enhanced TH implementation within ALEP, touching on what, where, who, when and how.

Given that there will be no extra funding for such an implementation, a very desirable feature is that it can be done within the existing LSGRAM+ resource allocations. It is agreed that this limits the level of ambition, however the core group in particular is interested in covering just basic phenomena such as dates, figures (i.e. numbers) and proper names.

In order to get a feel for what other phenomena groups might wish to be processed with a TH approach, it was decided that B.Music would circulate within the TH SIG a sample list of rules and phenomena covered by the PaTrans document handler in order that groups might compare this against their own data and give feedback regarding what phenomena they would prioritize for the ALEP TH, as well as general input on the new TH functionality.

Principles/features of the implementation of this functionality were discussed, i.e.:

It should be implemented with minimal changes to the existing system.

Reversibility of rule application is not an absolute requirement, as long as there is a consistent and defensible treatment of phenomena.

There should (naturally) be access to the user-defined patterns/regular expressions via ALEP's interface.

Brad Music of CST will come up with a preliminary design that attempts to integrate the points made at the meeting and the expected feedback.

An outstanding issue is where the integration with ALEP's TH should occur, e.g. before the existing TH, as part of Word Analysis or as a User Application.

To summarize, the second ALEP User Group workshop was, like the first workshop, a success. The discussions have shown that such a workshop is necessary in two respects: first, to

allow the ALEP community an insight into the actually running projects with ALEP and the opportunity to discuss different engineering approaches to the development of lingware and software under ALEP. Second, the presentations and discussions permit a kind of validation of the maturity of ALEP for real-life development, and therefore should also be of interest for companies involved in Language Industry. Unfortunately, the participation of industrial companies in such events is more or less zero.

If there should be a workshop at the end of the LRE-62101 project, which all of the participants of the second workshop would very much appreciate, then some members of the AUG have agreed to mobilize their industrial collaborators for a participation.

ASSOCIATION NEWS

International Association for Machine Translation
Third General Assembly
Hemicyle of the European Parliament
Thursday, 13 July 1995

Minutes

John S. White (Secretary)

The Third General Assembly of the IAMT was called to order by President Margaret King at 12:45 PM, with 60 in attendance. The provisional agenda was adopted, and also the minutes of the Second General Assembly.

In the Report of the President, Dr. King indicated that MT Summit V had been the main event in the activities of IAMT. There were over 400 attendees, including a large number of students.

The Secretary reported that he, John White, had been selected to complete the term of Scott Bennett, who had requested relief in light of his many professional commitments. He also reported that IAMT had been granted nonprofit status in the United States as of April 1995.

Scott Bennett, reporting for the Treasurer (Roberta Merchant), indicated that the IAMT was financially strong, although the exact figures were unavailable at the time of the General Assembly.

Makoto Nagao, representing the AAMT, reported that there were 190 individual members and 45-50 corporate members. The economic situation in the Asia-Pacific region had hindered growth in membership. The AAMT Journal was being published every four months in Japanese and English. There have been three workshops in Japan, where most of the members of AAMT are located. The proceedings of the workshops are available in Japanese.

Representing the AMTA, Muriel Vasconcellos reported that the organization has structured its meetings to occur in non-Summit years. There have been two workshops (1992 and 1994), and the proceedings of both of them have been published. Five special interest groups were developed that will feed into the next regional conference. AMTA will host the next MT Summit in San Diego in late 1997, in the western United States, though she is accepting proposals. The AMTA has 172 members; the elimination of associate memberships has reduced the numbers somewhat. The AMTA editor, Joseph Pentheroudakis, had resigned and had been replaced by David Clements.

Margaret King, reporting for EAMT, indicated that the Association has 63 individual members, five for-profit institutions, and five nonprofit institutions. There is a membership campaign under way. The most recent EAMT workshop was on the lexicon; the proceedings of a workshop on theory will be published this year. Professor King led a round of applause for John Hutchins, who has served as the editor of MT News

International and is the new President of EAMT.

Professor King presented the new officers of IAMT, explaining the relationship of regions, officer positions, and the MT Summit meetings. President-elect Muriel Vasconcellos of the AMTA became President, Professor Hozumi Tanaka of AAMT became President-elect.

Under New Business, Yorick Wilks asked about progress in the area of MT patents in relation to the resolution passed at the 2nd General Assembly in Kobe. Dr. Vasconcellos indicated that the problem has probably died.

A question was asked about the change in format of the MT News International. John Hutchins replied that the change was a function of change in publishers. Professor Hutchins made a call for columnists to contribute to the newsletter on a regular basis.

The date and place of the next MT Summit was established as coincident with MT Summit VI, to be held in San Diego in 1997.

There was discussion concerning whether the MT Summit and the Conference on Theoretical and Methodological Issues should continue to be held in the same time period and location. A straw poll of the attendees showed that 19 persons favored keeping the two together, 3 preferred having them separately, and 19 had no opinion. Other discussion concerned the content and orientation of the respective conferences.

After a vote of thanks for the organizers, hosts, and technicians who had worked on MT Summit V, the General Assembly was adjourned at 1320.

European Association of Machine Translation

The General Assembly of EAMT met during the MT Summit conference on the 11th July. Margaret King and Ian Johnson tendered their resignations as president and secretary respectively. Both were warmly thanked for their contributions in the foundation of EAMT. Other committee members also resigned, and the meeting elected the following committee:

President:	John Hutchins
Secretary:	Viggo Hansen
Treasurer:	Doris Marty-Albisser
Members:	Bente Maegaard
	Colin Brace
	Dmitri Theologitis
	Jörg Schütz (as EAMT newsletter editor)

The Assembly considered the state of the membership and the financial position and heard reports on the activities of EAMT during the past year, of which the major one has of course been the planning and organisation of the MT Summit itself.

The first meeting of the new committee took place on 13th July. Among the matters discussed were:

* a resolution to organize an EAMT workshop in conjunction with the TKE'96 conference in August 1996 [for more details see Conference Announcements]

* an agreement to launch a campaign to increase membership and to publicise the activities of EAMT.

* an agreement to start planning for the MT Summit in 2001, with bids for potential conference sites to be considered at a later time.

European Languages Resources Association

ELRA held its first General Assembly on 25th September 1995 in Luxembourg.

The main matters reported were:

* an announcement by Vincente Parajón Collada that the Commission had awarded ELRA a grant of 900,000 ECU over a three year period, starting on 1st October 1995

* membership of ELRA now stands at 66 (of whom 47 are fully paid up)

* the President announced that he had signed agreements with Dr Khalid Choukri for his undertaking the task of Chief Executive Officer of ELRA, starting on the 1st October 1995. Dr Choukri told the meeting that he would probably base his office in Paris.

* the Resolution to extend the area in which member organizations could be beased to the whole of Europe was approved, as was the Resolution enabling the Board to authorise, in exceptional circumstances, the membership of individual professionals.

The following officers were elected:

President:	Antonio Zampolli
Treasurer:	Thomas Schneider
Vice-Presidents:	Joseph Mariani,
	Norbert Kalfon,
	Angel Martin-Municio
Secretary:	Robin Bonthrone
Board members:	Louis Boves,
	George Caryannis,
	Giuseppe Castagneri,
	Christian Galinski,
	Harald Höge,
	Bente Maegaard
Chief Executive Officer:	Khalid Choukri
Commission Assessor:	José Soler

Registered office:

European Language Resources Association
c/o CL International
46 Grand rue
L-1660 Luxembourg
Tel: +352 46 91 60; Fax: +352 46 91 61

ELRA has now published for its members the first draft of its survey of existing and planned European language resources. Researched under the aegis of the European Commission through three of its major resource projects, RELATOR, SpeechDat and PAROLE, the survey has been edited by Elizabeth Hunkelman (DFKI), with help in formal presentation from Colin Brace. The information covers resources in the spoken and written fields, with relatively little space devoted to terminology; it is intended that this will be rectified in future editions. The ultimate aim of the survey is to present as complete a spectrum as possible of current and future language resources in Europe. The subsequent role of ELRA will be threefold: in that of broker, putting the user in touch with the creator or owner; of validator, defining and applying the standards necessary to transform the information into a marketable form; and of distributor, disseminating those data available to a broader public. Subsequent editions of the survey will be published as and when more information becomes available to update, correct and complete the material.

Asia-Pacific Association for Machine Translation

[From AAMT Journal no.11, June 1995 (unedited, except spelling)]

The Fifth AAMT General Members' Assembly
Excerpt of the Agenda

A. 1994 business report

1. Research and study activities

It is necessary to enrich the information transmission infrastructure attempting the widespread and enlightenment of machine translation application. Therefore, AAMT held the research and study meetings up to 21 times in total: the work groups for sentence evaluation and market discussed those relevant problems, and they estimated future conditions as well as completed an illustrative research of the actual MT application conditions.

(1) Sentence evaluation work group

Aiming at how they will advise each company to improve their MT systems, the sentence evaluation work group extracted and accumulated those sentences a user may find difficult to translate. It examined the pre-editing and rewriting of the same sentences based on the translation result of each company's systems. Moreover, the work group processed those pre-edited samples once again with MT. Also, native English speakers examined and evaluated the latter MT results, too.

(2) Market forecast work group

(i) The machine translation doesn't stop at merely translation work. It pays attention to the function to collect and communicate the information in foreign languages. This work group discussed the means which the machine translation contributes for the translation productivity as well as upgrading productivity of information in the job flow.

(ii) The work group learned the basic knowledge of the relation between the office automation (OA) and the machine translation aggressively lectured by the professional in that field. The lecture enrolled members of other work groups.

(iii) It analyzed the current of diversification of the translation softwares encouraged by the hardware price decline. It examined the trend.

(3) Application ware study activity

(i) Introduction of the MT softwares for Windows made the operability of the MT improve notably. All kinds of the functions recently has come to reflect the views of the user, too. To ascertain this trend, the members exchanged market information collected from each company's users.

(ii) Lately, the situation surrounding MT has changed suddenly, which demands the work group to look into progressive study subjects.

(iii) Separate from the above, the work group prepared a text for the users' workshops.

2. Projects to promote and upgrade MT application

Based on the result of the research activities experienced, AAMT held a seminar and users' workshops. It improved the popularity of MT systems while implementing the promotion of MT.

(1) MT Seminar

AAMT held a study report meeting for the public. At that occasion, study activities of fiscal year 1993 was publicised to the public in general.

Date and time : June 22nd, 1994. Venue: Tokyo.

Themes: (i) Study reports of the work groups

(ii) Development of EB-MT system

(iii) Three case studies of MT users.

(iv) Lecture: "Trend of internationalization and translation"

(2) Workshops for user training

Date and time : November 25th, 1994. Venue: Nagoya. Co-sponsored by the Nagoya Chamber of

Commerce.

Themes: (i) Outline of MT
(ii) Function outline of each company's system
(iii) The user participated in the demonstrations.

Demonstrators: four MT vendors. Total of twenty end processors were used.

(3) Co-sponsored seminars

(i) TC Symposium - AAMT session

Date and time : August 26th, 1994. Place : Tokyo.

Themes: (i) Outline of MT
(ii) Pre-editing
(iii) Post-editing

Technique of later processing

(ii) Translation Day Symposium - AAMT Session
Sponsored by Japan Translation Federation.

Date: October 1st, 1994. Place: Osaka.

Themes: (i) How MT was born.
(ii) Machine translation dictionaries

(4) Site visit to translation enterprizes

Date and time : April 6th, 1994. Place: Hamamatsu City, near Nagoya

Themes: Document proofreading software and upgrading MT efficiency

(5) Lectures offered by professionals

(i) September 30th, 1994: "Industrial translation revolutionalized" by Mr. Takeshi Narumi.

(ii) November 30th, 1994: "OA and office work productivity" by Mr. Yoichiro Suzuki.

(iii) December 16th 1994. "My experience of translating technical documents with PC MT" by Mr. Tanehiro Tatuta

(iv) January 24th, 1995: "Productivity compared between MT and Human Translation (HT)" by Mr. Shinya Amano

(v) February 15th, 1995: "Upgrading translation efficiency with MT" by Mr. Terumasa Ehara

3. Co-sponsored Projects

AAMT extended cooperation with the groups of the MT and the related circle. It communicated and collected information, and made the results available for its members through the AAMT Journal and so on.

(1) August, 1994, at Tokyo. "Coling'94": tutorial sessions. Cooperated for the Public Relations.

(2) October 21st, 1994, at Tokyo. "Translation Fair '95" (Sponsored by Japan Translator's Association). Cooperated for the Public Relations.

4. Cooperated Projects

AAMT communicated with its affiliates such as JTF, JTA, JPDEC, JEIDA, CICC and EDR and it implemented various research, information exchange on various study results and publicised events.

5. Publication of AAMT Journal and PR activities (Editorial committee)

Product information, user case studies, invited lectures and introduction of related technologies: AAMT collected opinions and so on aggressively and provided news and resources to its membership. Authorized IAMT's journal "MT News International" reprint/excerpt information from AAMT Journal and provided them worldwide.

(i) AAMT journal (Japanese/English editions)

Introduction of new products: Lectures for utilizing MT; Country report; Symposium report; Topics and so on.

(ii) MT Newsletter No.8-10 : Machine translation study of the world; Technology development

trend; Event announcement.

(iii) Study Report. 94 Edition: System Utilization Technique Work Group; System Evaluation Work Group: MT Market Forecast Work Group.

(iv) MT Summit. Distribution of Proceedings.

6. International Cooperation

AAMT cooperated with the affiliating institutions/bodies abroad and domestic, and it tried for the extensive information alternation.

(1) with IAMT: AAMT publicized MTNL in its region and exchanged the international symposium information.

(2) with Center for International Cooperation for Computation (CICC): AAMT collected Asia related information about the information processing. It provided event information and so on on the bulletin for the PR. Also, for further promoting MT, it is providing CICC library in 29 overseas countries with AAMT Journal.

Planned Projects for fiscal year 1995.

1. Projects

(1) To further the Research and development activities. The information dispatch infrastructure as the basis for the popularization and promotion of the machine translation system should be maintained and reinforced . It intensifies and develops so far the investigation and research of machine translation aggressively through its Committees and work groups (WGs).

Moreover, as the Association also progresses its activities into the fifth year this year, the nomination and promotion of new research theme will be executed.

(2) Promotion and popularization of MT. The Association cooperates with related groups and institutions and endeavors to promote activities: to improve the eminence of the machine translation; to achieve settled and efficient application of MT among circles at various levels; to recruit potential users; to expand wider and further usage, and for efficient application of machine translation.

(3) Maintenance and reinforcement of the Association's management structure. The maintenance of the Association's management structure calls for the expansion of the membership, thus recruitment and the participation of further MT user is aimed at. It is indispensable that the extended membership, from not only among the industry related to the translation, will reinforce the Association as well as stabilize its management.

2. Particular Projects

(1) Holding of workshops, lectures, and visits to MT related industries.

* For research and development activity for MT, it will investigate the technological trend, the market trend, and also continue the research for corpora evaluation.

* It will visit sites in translation industry, and the case studies how each site utilizes/employs MT will be investigated.

* The seminar to promote the result of each research will be held. The text is edited by the workgroups and committee members.

* To improve recognition related to machine translation, the lecture and workshops will be held.

(2) Publication of newsletters and reports.

* AAMT Journal publications will be continued.

* It will print the result of researches in its Journals so that R&D achievement will be promoted.

* Worldwide MT information is introduced through the issues of "MT News International", edited by IAMT. It will take responsibility to distribute those issues in its district.

(3) The MT training and education courses

* In cooperation with the affiliated institutions and related groups, it will hold on-the-job training sessions for MT users in various places.

* It will participate in those seminars and so forth in cooperation with other groups and institutions.

(4) Technological development concerning MT

* It will try to exchange the technological, technical and other information based on the result of its various research.

* The exchange of information on MT is aggressively made by supporting and cooperating with academic and other bodies.

(5) Research activities

* For research schemes and formats for system evaluation studies: those for corpora evaluation and other themes will be examined the counter measure.

(6) MT utilization guideline

* MT utilization technique will be investigated for research, and on-the-job training handbook will be edited.

(7) Information retrieval for technological investigation for MT by other institutions, MT related data and documents/publications

* It will feed back its membership through its Journal and other medium such collected/accumulated technical documents as released from those various research organizations, related academies, affiliates and government and municipal offices.

* The experts of each business/research domain will be invited as the lecturer for the work group/committee meetings, where information is collected and exchanged.

(8) The promotion of the academic research and development

* It is aimed at through symposia and the seminars of the investigation and promotion of those research.

* It will support and sponsor the symposia, seminar and other events of related academic institutions as well as affiliated institutions.

(9) Cooperation maintained with related groups and various bodies/institutions

* Other than those domestic related groups, cooperating with various groups in those countries/regions in the Asia-Pacific district, it will offer and collect information for them.

* Cooperation with overseas groups affiliated to the IAMT is reinforced and information exchange and so forth are activated.

(10) Additional projects

* It will examine and execute projects aggressively, based on the demand and the proposal of its membership aiming at the popularization and promotion of MT.

Bill 2 - Financial report for 1994.

2-1: Closing of amounts of 94 terms and 1995 fiscal budgets.

PART OF INCOME.

[ITEMS:	1994 fiscal budgets/	results/	1995 fiscal budgets/]
Membership Fee.	15,000,000	12,455,000	12,000,000
Admission fee.	30,000	22,000	30,000
Miscellaneous incomes.	500,000	3,222,411	600,000

Interest receivables.	150,000	186,130	150,000
	(150,000)	(186,130)	(150,000)
Total income this season	15,680,000	15,885,541	12,780,000
Balance brought over from the last account.			
	15,737,292	15,737,292	16,131,120
Income total.	31,417,292	31,622,833	28,911,120

PART OF EXPENSE.

[ITEMS: 1994	fiscal budgets/	results/	1995 fiscal budgets/]
Conference meeting	2,430,000	1,240,064	2,150,000
Membership	300,000	239,000	250,000
Printing	2,400,000	1,499,239	2,000,000
Translation fees.	400,000	548,658	650,000
Clerical works.	8,450,000	6,832,318	8,360,000
General expense	300,000	257,500	600,000
Office rental fee.	4,600,000	4,573,200	2,500,000
Commission.	20,000	10,854	20,000
Guarantee allow. depletion.	280,000	290,880	150,000
Reserve expense.	12,237,292	0	12,231,120
Expense meter of this period.			
	31,417,292	15,491,713	28,911,120
(Revenue and Expenditure difference at this season)			
	393,828	0	
Total revenue and expenditure difference.		16,131,120	

AMTA Coordinates Initiative for Standardized MTAPI

One of the challenges that software developers face if they want their products to offer integrated access to machine translation is that they have to write different code for each MT system. This discourages them from incorporating translation as a standard feature.

Over the last few months several MT developers have been working to standardize the calls to their software so that a developer can access a single MT applications programming interface (MTAPI) that will allow any "client" program to call on any MT "server" that uses the standard. Thus it will enable many different programs, including word processors, to offer the functionality of MT to solve the needs of users in an international marketplace.

Mindful of the importance of this issue, AMTA took the initiative last May to establish a subcommittee of its Special Interest Group on Standardization to coordinate developments in this area.

In August of this year a draft developer's kit for an MTAPI was demonstrated. It is based on OLE Automation, the Microsoft standard for interprogram communication. By using OLE, the MTAPI takes advantage of the features in Windows that permit applications to be linked, or make calls to, other applications. In this way, users can add translation capabilities to their favorite word processor, database, spreadsheet or on-line communications software. "Later," said Timothy Meekhof, Chief Scientist at Globalink, who worked on the proposed draft, "the MTAPI will be expanded to other operating systems besides Windows."

The demonstration kit consists of an MTAPI-compliant Pig Latin translator that is called by a simple Visual Basic program. Visual Basic was chosen for the sample application since it inherently supports OLE, although other languages such as C and C++ can support OLE with some work.

While a simple Pig Latin translator is hardly exciting news, the fact that it is MTAPI-compliant is. "Developers can now integrate translation software without tying their code to a specific language, vendor or version." said Michael Tacelosky of Globalink, chair of the MTAPI Subcommittee. Denis Gachot, President of Systran, envisions the day when "the AMTA logo on the box will tell the public that the software inside complies with our industry standard."

Initiated by Globalink, Inc., the proposed MTAPI has received the support of other MT vendors, including Systran and Logos. According to Sean Ivory of Systran, "the MTAPI should be implemented ... by early 1996." Mark Miller of Globalware, publisher of XL8, adds his "enthusiastic support," while emphasizing his desire to see the standard expanded to include features necessary for "translation for information dissemination." Such features would include the ability to recognize text formatting as information to be accounted for in translation. Meekhof indicates that the first version of the MTAPI will not include such a capability because it would make the standard "too complicated and reduce the number of programs that could use it."

Developers, both "clients" and "servers," that wish to participate in this initiative should contact Michael Tacelosky, Chair, at tac@globalink.com.

PRODUCTS and SYSTEMS

English-Japanese translation support software "Trans-supporter EJ" by SANYO Information Business Co.

[From AAMT Journal, June 1995 (unedited, except spelling)]

Introduction

The development of translation support software has entered its second stage.

With the spread of networks such as internet or commercial PC network, it is becoming more common to send electronic mail overseas or to access database in another country.

It is more likely that an imported software will have manuals written in language other than user's native language. Whether we like it or not, we will be forced to deal with a foreign language in our everyday situations.

As the performance of a personal computer improves and as its operating system becomes more user friendly, the machine translation technology has become a familiar technology to us. Translating an electronic mail or a news article on a personal computer has become a simple matter.

Sanyo has been working on the improvement of the machine translation technology since we announced the development of Japanese-English translator word processor "Trans word processor" in February 1987.

This time, we are announcing "Trans-Supporter EJ", an English-Japanese translation assistant for Windows and pen-type OCR.

2. Characteristics

[1] English-Japanese Translation assistant "Trans-Supporter EJ" for Windows

(1) Translation engine and dictionary

By adopting multi-phase transfer method of translation, we have achieved significant improvement in the analysis ability, especially for embedded and long sentences. Trans-Supporter EJ also has a database of translation templates for translating stereotype expressions.

The system dictionary has 80,000 words and it is possible to maintain different translation environment for each user by preferable translation setting and user dictionary registration. The user dictionary can then be saved to a disk for later use.

Also available are dictionaries for technical fields (electronics, machinery, economics, and medicine).

We are proposing a new form of translation assistance. For the ease of operation, Microsoft Word(5.0/6.0) will be used for word processing. Whenever it becomes necessary to translate, Trans-Supporter will be called up to assist translation.

Thus, a user will be able to utilize the useful tools of a word processor for document creation when translating using the TransSupporter.

There are two ways to translate file(s) using TransSupporter.

When working in interactive mode, each sentence and its translation are successively checked in a post-editing window (Figure 1) after the translation.

When translating multiple files, specified files are automatically translated and stored in batch translation mode.

(2) Dictionary reference (referring to built-in translation dictionary)

Even if there is the word you do not know, you can refer to the dictionary by clicking the word only.

[2] pen-type OCR (pen-type OCR is used to input text)

You can use to read easily the printing type in Japanese/English as supporting machine for inputting the printing type in the printing matter such as the catalog and the newspaper.

a) It supports both Japanese and English characters.

b) It can recognize both lateral and length-wise writings in Japanese.

c) It is equipped with simple determining mechanism from recognized candidates.

3. Specification

[1] Translation assistant software "TransSupporter EJ"

System dictionary : 80000 words

Technical dictionary (optional)

4 fields (electronics, machinery, economics, medicine)

User dictionary: (limited only by available HD area)

Operating environment:

requires Japanese Window 3.1 and Microsoft Word 5.0 or higher

MPU: i486SX or higher

memory: 8 MB minimum

HD: 18 MB

[2] recognition method of pen-type OCR : by hand

recognition rate: over 99% (according to Sanyo guideline)
recognition speed: 4 characters / second
PC-98 or AT compatible DOS/V personal computers

[For more information]
SANYO Information Business Corporation, 2-7-25 Edobori, Nisi-ku,Osaka-shi 550, Osaka, Japan. Tel 81-6-443-5144

Japanese-Korean Machine Translation System(HICOM-MT)
HItachi COMmunication network – Machine Translation
by Hitachi Information Network, Ltd.

[From AAMT Journal, June 1995 (unedited, except spelling)]

1. Introduction
Recently, economic and social exchanges between Japan and Korea of which economy has rapidly grown up are brisk. We had developed a Japanese-Korean Machine Translation system (HICOM-MT) in last September. The system has two main purposes – the first: going smoothly to communicate between Japan and Korea, the second: raising the efficiency of translation business. And, in March '95, we had also developed a Korean-Japanese Machine Translation system making pair with this system.

2. Features
(1) Use of the optimum translation method for Japanese and Korean languages
Japanese is very similar to Korean in concerning to the followings.
(a) Structure of a sentence
(b) The role of postpositions, which is so-called a jyoshi in Japanese.
(c) A inflective predicate
A syntactic direct method which exchange words, phrases and clauses in a Japanese sentence was used because of the above similarity.
(2) A user can select the appropriate Korean equivalence to a Japanese word
If there are some Korean equivalences to a Japanese word in a sentence, a user can describe a rule for selecting the appropriate Korean equivalences. By means of the feature, the translation accuracy of a sentence raises more. As he can easily describe the rule, he can soon get a translate useful in various ways.
(3) The MT translates up to 66,000 characters per hour with I486DX2 66MHz.
(4) Selection of kinds of Korean characters in a translation result
(a) Hangul character only
(b) Hangul mixed with kanji
(c) Hangul mixed with alphabet and kanji
(5) System dictionary approximates to 100,000 words.
(6) Type of translation styles
the MT has the following two types.
(a) Translation in interactive style

(b) Translation in batch style

A user can translate up to 32 Japanese documents in succession. The document may be held on DOS file. It is useful in translating a large quantity of documents.

(7) A user can print a source text and a target text in combination with each on A4 size paper.

(8) Maintenance of system dictionary

The system dictionary and a user dictionaries can be maintained easily.

3. Specifications

Hardware: machine executable for DOS/V with VGA

Memory required : 3.8MB or more

OS : MS-DOS 5.0J/V or more

Printer applicable: Page printer with ESC/P

Font for printing Japanese : must be involved in your printer or MS-DOS

Input Korean characters : use Roman-ji method or Two ber method which is Korean standards.

4. Price: HICOM-MT 298,000 Yen

For further information: Network Solution Division, Hitachi Information Network, Ltd. IK Building, Omori-kita 1-18-2, Ohta-ku, Tokyo 143 Japan Tel: +81 3 5493 4815 Fax: +81 3 5493 8472

Windows-based Translation Tool "Korya Eiwa!" Catena Corp.

[From AAMT Journal no.10, March 1995 (unedited)]

1. From Translation System to Translation Tool

The original task of the machine translation system was simply, 'to translate'. Then, one after another, features which might prove useful during operation were added to the systems. This resulted in option-filled, giant translation systems with feature-rich menus being regularly developed.

A translation specialist may well be delighted to have this kind of multi-feature system, being able to select only those features that are truly useful to him, and using this personally customized system.

However, for the general user, this abundance of features would seldom if ever prove advantageous. For this kind of user, the ability to check the system's dictionary for, or translate unknown English words with little effort is much more essential. Likewise, for a great number of these users, things like pre-editing to improve the translation results are quite unnecessary and deemed superfluous.

Moreover, elaborate, multi-function systems may give the impression of being too complicated and even rather difficult to approach, and this may contribute to the slow growth of the machine translation market, despite the potential demand.

At Catena, we considered these facts and realized that a more easy-to-use translation system -- a translation tool -- was needed. This would be a tool to use when looking up words or translating small amounts of English encountered in game instructions, E-mail, etc. Thus did the Windows-based English-to-Japanese translation system "Korya Eiwa!" come into being.

We pared the system down to the absolute minimum number of features. We even cut out some features that are usually considered standard in translation systems. For example, realizing that users of this system would primarily be translating, reading and discarding the translations, there is no function allowing the user to save the translated material into files. (Of course, because it is a Windows-based system, one can copy the material into a different application and save it from there.)

In place of these generally unessential features, we have added new features, such as the "Clipboard Supervisor" (explained later) and have set the suggested retail price at an affordable 9,800 yen.

And thus, this first attempt to create a tool of this nature was a step into the unknown, but thanks to the many people who are using it, "Korya Eiwa!" seems to have been given a very kind reception.

2. A Professional Translation Engine

With the low price and the downsizing from system to tool, one might be tempted to think that the translation engine could be in some way substandard. However, this is not the case. In fact, the engine sued in "Korya Eiwa!" is the same that is in our original translation system.

Catena has developed a high-speed, highly-accurate and highly-acclaimed UNIX-based translation system called "STAR". We have placed the "Star" engine as is into "Korya Eiwa!"

This is the same technique that was used in creating the Macintosh-based "The translator" translation system series, which also has obtained an excellent reputation. And, because the "STAR" engine is compact and has multi-platform interface capabilities, the cost of developing this new tool was kept quite low.

The system dictionary is larger, too, than ever before. Our latest product improvement boosted "The Translator 2.5" dictionary to 62,000 entries, but "Korya Eiwa!" boasts a dictionary of 67,000 entries. Moreover, in addition to increasing the size of the dictionary, we have made changes in it to further improve translation accuracy.

3. Translation and Dictionary Access are Kept Separate

Translating and checking the dictionary are independent applications, so the user will find that it is quite convenient to primarily use the dictionary to look up words. (Of course, though, even while translating, the dictionary can be automatically accessed with the dictionary tool, so complicated procedures are unnecessary.)

4. The Clipboard Supervisor Feature

This feature allows the user to copy, for example, any English that is not understood from games, E-mail, News, etc. to a clipboard where it is automatically looked up or translated. This feature is extremely easy to use. The user begins by simply clicking the "Clipboard Supervisor" feature on the menu and proceeds to play games or read E-mail. Then, when some words that the user does not understand appear, they can be copied and sent to the clipboard for help.

Having this feature is like having a professional translator and linguist sitting quietly on your desktop waiting to appear when you need them. Use it once, and you will realise just how helpful this is.

Of course, English text may be read in from a disk or other file and translated in the traditional translation system method.

5. The Translation Tool

As was previously stated, anything that would tend to make the system seem too complicated or difficult to approach must be avoided, so while the main features of this translation tool will now be introduced, we refrain from presenting them all at this time.

* Right/Left or Top/Bottom Alignment

For parallel translations, a right/left alignment seems to be quite popular these days, but we have decided to put both right/left and top/bottom options into this system, there seeming to be no reason to limit it to just one or the other.

* A Five Page Translation Pad

Up to five translation screens may be used at once. (One of these is used by the Clipboard Supervisor.) These screens are independent of each other, and are saved automatically when the translation tool is turned off. (This automatic saving of screens in a small way makes up for the previously mentioned lack of a file saving feature.) Each pad will hold English and translated materials of up to 32KB.

* Optional Whole or Sentence-by-Sentence Translation

The user may choose either to translate one part of the text or the whole text, and whether to translate that material all at once or sentence by sentence.

* Optional Partial English Mode

Words not entered in the dictionary will, of course, be put into the Japanese translation in the original form; but, if desired, the user may elect to have even nouns and verbs that are in the dictionary left in English.

This is a feature that was originally in the "STAR" system and was placed into "The Translator" system also. It is useful especially in documents with large amounts of special terms or professional jargon. We add this feature because we have been told that for certain people, reading these words in English is, strangely enough, much easier than reading them in Japanese.

* Post-editing and Adding Words to the Dictionary

These may seem to fit the image of the traditional translation system more than that of a translation tool. With these features the user can adjust the translated word to better fit the meaning of the original English word and to change the translated word's position in the sentence.

The dictionary can be enlarged by adding new words more likely to be satisfied with the results.

While it is naturally the ideal that the quality of machine translation rise and the user be provided with high-quality translations, until that day comes, we hope that "Korya Eiwa!" will lead the way in the popularization of machine translation.

For further information, contact:

Catena Corp. (attn Mr Matsuda, Customer Service). Fax: +81-423-38-1079

The words "Korya Eiwa!" means "This is English-to-Japanese [translation]", and sound much the same as the Japanese for "Oh, this is great!"

"A polyglot photocopier"
Autoscribe out of Japangloss

If most of the general public are asked how they think a translating machine should work they would probably say that it should be a device like "a photocopier that can translate documents at the push of a button: in goes Japanese, out comes English." This is precisely how **Newsweek** (7 August 1995) describes the prototype MT system called Autoscribe developed by Kevin Knight and his colleagues at the Information Sciences Institute of the University of Southern California. It goes on to describe it as "something of a hardware mutation -- a color scanner housed in the skeleton of an old copier. The scanner is linked to a SUN workstation, which runs the software needed to translate the document." It is gratifying that the brief report injects some words of caution: "A slipshod translation can take 4 minutes per page; true accuracy requires 15. But foreign-language students sweating over that essay shouldn't hold their breath. Autoscribe could cost more than $10,000 and won't be in stores for a few years."

More accurate details are contained in a report in the university's own newspaper under the title "A polyglot photocopier". The prototype uses a commercially available Japanese character recognizer from MRJ Inc. of Fairfax, Va., whose 6 percent error rate has been reduced to less than 2 percent by an ISI student Kenji Yamada. The translation program is based on the Japanese-English MT research at ISI as part of the Pangloss project under the direction of Eduard Hovy. This interlingua knowledge-based system (Japangloss) is designed for unrestricted newspaper articles. It includes components typical of KBMT systems: a syntactic parser, a semantic analyzer to produce candidate interlingual representations, a semantic ranker to eliminate meaningless interpretations, a generator for producing English output, and a range of knowledge databases: Japanese grammar and semantic rules, inference procedures, concept constraints, English grammar rules, Japanese-to-English lexicon, etc. However, in order to cover unrestricted text Japangloss has integrated statistical methods for the acquisition of lexical and grammatical information from online dictionaries and from texts themselves. For example, the project has derived bigram and trigram transition probabilities for English text from a large collection of 46 million words of the Wall Street Journal. Similarly, Japangloss uses part-of-speech bigram probabilities to improve parsing performance.

The report emphasises that "The translations that Knight's program currently can output would hardly pass muster as clear (or even comprehensible) English. But less-than-perfect machine translation can nevertheless be useful to analysts needing to monitor large volumes of foreign-language text." And it quotes Knight's belief that "improvements in translation and character recognition software should make a commercial version of his model a functioning reality within five years."

[Extracts from University of Southern California Chronicle, August 28, 1995 (vol.15, no.1), p.1-2.] For more information: Kevin Knight, Information Sciences Institute, 4676 Admiralty Way, Suite 1001, Marina del Rey, Ca. 90292-6695; fax: +1 310/823-6714.

CompuServe Launches Document Translation Service

[Press release]
The CompuServe Information Service now offers a way to translate letters, articles, mail messages

or business documents. The CompuServe Document Translation Service offers electronic translation between English and French, English and German or English and Spanish. Members can choose one of two levels of service: unedited machine translation or translation that includes human post-editing for a greater level of accuracy.

The service offers a front-end with menus in each of the four languages, accessed by any of four GO commands (GO TRANSLATE, GO TRADUIRE, GO TRADUCIR or GO ÜBERSETZEN). The cost for unedited machine translation is three cents per word. For translation that includes human editing, the cost is 10 cents per word, with a $50 minimum. To use the service, members must submit electronically files in either Text or Rich Text (RTF) formats. Translated documents are delivered to the member's CompuServe mailbox. Unedited translations are often returned within minutes.

Human translation editing services are provided by Linguistic Systems, Inc. of Cambridge, Mass. In August 1994, CompuServe introduced machine translation, first used to translate electronic communication among members on-line. CompuServe is the world's on-line global information service, with members in more than 150 countries. For more information, contact Carrie Reber or Michelle Moran at CompuServe Incorporated, 614/538-4092 or 614/538-3497.

Globalink Forms Latin American Partnership with Digital Equipment Corporation

[Press release]
Globalink, Inc. has negotiated a strategic agreement that will provide Digital's business customers throughout Latin America and the Caribbean with the first PC equipped to work in both Spanish and English.

In the agreement reached with Digital Latin America Caribbean Territory (LACT), Globalink has established a broad relationship with Digital's subsidiaries and distributors in Latin America and the Caribbean. Under the agreement, Globalink and Digital will provide PC customers with preloaded or bundled versions of Globalink's Spanish Power Translatorr software, a mid-level business solution, and its professional product, Spanish Power Translatorr Professional which includes industry specific subject dictionaries.

"For the first time Latin American business customers will be able to purchase a premium quality PC that will enable them to work in both Spanish and English. Not only does this offer Digital a competitive advantage in the retail PC market, but it also provides its customers with an advantage – the ability to conduct business globally without concern to language barriers," said Digital LACT's general manager, Andres Carvallo. Jim Lewis, Globalink's President, agreed, "Globalink offers a resource that will allow even medium to small sized businesses to compete beyond their nation's borders. Digital is particularly well qualified to deliver that resource to Spanish speaking customers in Latin America and French speaking customers in the Caribbean.

The agreement will first be implemented in Venezuela, where Digital's PCBU general manager Alberto Castillo said, "The fact that Globalink's language translation software compliments our word processing application is very attractive. What really impressed us, however, was the Petroleum & Mining subject dictionary. This feature provides our country's booming oil industry with something that uniquely meets their needs."

Globalink to Provide Netscape Clients with On-Line Language Translation

[Press release]
Globalink, Inc. announced on September 18, 1995 its participation in Netscape Communications Corporation's Development Partners Program. Globalink's software will be integrated with Netscape NavigatorT 2.0, allowing users to translate foreign languages while on-line.

"Today's Internet activity is conducted in many languages. There is a wealth of information that just isn't accessed because there is no convenient way to translate it. The ability to obtain a document from the Internet in a foreign language and use Globalink's software to receive a gisted [sic] translation, offers incredible possibilities for global communication. We are excited that, like Globalink, Netscape also has the vision to connect people around the world in spite of language barriers," said Globalink's president Jim Lewis.

As a member of the Netscape Development Partners Program, Globalink will work with Netscape's open software platform to create applications for enterprise networks and the Internet. Globalink will provide translation capability for French, German, Spanish and Italian, to and from English, for Netscape Navigator with Globalink's next generation of translation technology, called "Barcelona." Users wanting to translate Internet information while viewing it through Netscape's software will simply access the translation feature and obtain a draft translation, suitable for comprehension of the source text, without disrupting their on-line search.

News from Lingvistica '93 Co

Michael Blekhman

PARS for Windows

We have developed a new version of our English-Russian-English MT system, PARS 3.1, which works as a Windows application and is WinWord compatible. The user may start PARS 3.1 either from Windows or from WinWord. In the latter case, starting MS Word 2.0 or 6.0 causes appearance of the PARS icon in the WinWord main menu, so the user may have a text translated into/from Russian without exiting the text processor. Translation direction as well as PARS dictionaries to be used in the translation session are set up by the user.

In the end of the session, a new window is opened in WinWord, displaying the target text. All the polysemantic words are marked as such, and the user may select a more appropriate translation variant and have it pasted into the text to substitute the initial variant. Proper names are considered polysemantic, and their transliterations appear in the list of variants.

Another option is dictionary extending by means of entering "new" words directly from the WinWord text.

The source text format is fully preserved in the target text, including tables and fonts.

PARS is supplied with a Cyrillic screen/keyboard driver and a set of printer fonts. The User's Guide as well as the screen Helps are in English and in Russian.

The system makes use of a unique set of bidirectional dictionaries including such topic areas as economy, machine building, aerospace industry, computers, microelectronics, mining, medicine, ecology, etc. The total number of word entries is 300,000 in each part (English-Russian and Russian-English).

Lingvistica '93 Co and ETS Ltd (Russia) have launched a joint project that includes creating a set of polytechnical dictionaries for PARS on the basis of the POLYGLOSSUM dictionaries by ETS. In the framework of the project, the PARS dictionaries will include about 1 million word entries in each part. The work will be completed within several months according to the dictionary compilation technology described in MTNI#11.

PARS 3.3/DOS and PARS 3.1/Windows are supplied in one package. Both versions make use of the same set of dictionaries. PARS goes either on floppy disks or one a CD ROM, in the latter case - with the POLYGLOSSUM dictionary support software.

The DUE System

Our company, together with the Center for Information Technologies of the Kiev Palace of Youth, have developed the DUE 1.0 English-Ukrainian-English dictionary support system. It is a Windows application, which provides:

- translating words and phrases between English and Ukrainian; if a word is not found in the dictionary, the most similar ones are displayed for further translation;
- checking separate words and full texts for spelling mistakes, which is especially important for those writing texts in Ukrainian; the system takes into account morphological characteristics of the Ukrainian words, such as declension and conjugation.

DUE is supplied with an English-Ukrainian bidirectional dictionary of 30,000 general usgae, socio-political and business terms, and a similar dictionary of 17,000 terms on ecology.

Further developments include new dictionaries and a Ukrainian grammar checker.

Fur further details, please contact Michael Blekhman: 94a Prospekt Gagarina, apt.111, Kharkov 310140, Ukraine. Tel.: (0572) 27-71-35.

Copernicus and the ALEP System

ALEP (The Advanced Language Engineering Platform) is an initiative of the European Commission to provide the natural language research and engineering community in Europe with a versatile and flexible general purpose research and development environment.

ALEP is a graphical, object-oriented system running under SunOS 4 or Solaris on Sparc machines. It includes a number of tools for grammar and lexicon development, debugging, text handling, and linguistic processing. Although ALEP is in principle formalism-independent, it is delivered complete with a 'lean' formalism implementation and demonstration lingware. (Detailed hardware and software requirements for running ALEP are described below)

It has recently become possible for the CEC to release the ALEP software within Eastern Europe. Some funding may also be available under the Copernicus programme to assist in the deployment of the system.

In the first instance, the CEC aims to provide the ALEP software, together with the requisite supporting software and licences (principally Prolog by Bim, ClauseDB, MOTIF) to a maximum of 10 user sites in Central & Eastern Europe, free of charge.

Towards the end of the year, the CEC also intends to hold an ALEP training course in Luxembourg, for up to 15 participants. It is hoped that travel and subsistence expenses for attendees from Central and Eastern Europe can be reimbursed. A limited number of on-site visits by ALEP support staff may also be undertaken.

Organisations who are interested in obtaining the ALEP software and participating in the ALEP course are invited to contact the CEC's support contractor Cray Systems at the address below, stating the nature of their involvement in language engineering and the uses to which the ALEP system might be put. Further information on the ALEP system may be obtained from: Cray Systems, 151 Rue des Muguets, L-2167 Luxembourg (Tel: +352 427744; Fax: +352 439594; Email: alep-support@cray-systems.lu.)

Telematics Programme: call for proposals

The Language Engineering Sector of the Telematics Application Programme is launching its third call for proposals on 15 September 1995. Closing date 15 is January 1996. Priority assistance will be given to:

* Innovative pilot application projects exhibiting a broad European scope, a significant user dimension, and a clear deployment and exploitation potential. These projects will concentrate on the integration, validation and deployment of well-mastered language technologies. Proposals addressing the needs of elderly and disabled people are welcomed.

* Leading-edge pilot application projects contributing to the emergence of an open Information Society through research and demonstration efforts aimed at countering information under- and overflow resulting from the growing availability of multilingual multimedia content over broadband and mobile communication networks.

These projects will foster the consolidation of novel language technologies and their integration into multimedia information and communication products and services. RTD will aim at the next generation of telematics applications, and will feature focused, goal-oriented research efforts.

There is a specific effort in this Call to encourage proposals falling within the scope of and contributing to Global G7 Projects in the area of multimedia information access and management. These proposals are expected to bring together researchers, information providers and system integrators, and to co-operate closely with on-going and planned European and international initiatives in the field.

* Horizontal and support actions addressing: pre-normative standards and guidelines; assessment of language components to be integrated into larger telematics products and service; European conferences and exhibitions; and user surveys in several countries and socio-economic sectors.

Documents relative to this call can be obtained from the Telematics Help Desks (see below) or from the WWW at URL http://www.echo.lu/. Outline proposals may be submitted using the forms contained in the TELEMATICS Information Package, preferably by facsimile, as soon as possible but no later than 30 November 1995, to the following address: European Commission, DG XIII-E-5 LE Office, Batiment Jean Monnet (B4-002), L-2920 Luxembourg (Fax: +352 4301 34999)

All specific inquiries regarding the LE Sector can also be obtained from this office. For all general inquiries regarding the TELEMATICS APPLICATIONS PROGRAMME, please contact: DGXIII-C Rue de la Loi 200 (BU29, 4/41), B-1160 Brussels (Fax: +32 2 295 2354; E-mail: telematics@dg13.cec.be)

CORPORA

WordNet Version 1.5 now available

WordNet is an online lexical reference system. Word forms in WordNet are represented in their familiar orthography; word meanings are represented by synonym sets (synset) – lists of synonymous word forms that are interchangeable in some context. Two kinds of relations are recognized: lexical and semantic. Lexical relations hold between word forms; semantic relations hold between word meanings.

To learn more about WordNet, read "Five Papers on WordNet", available via anonymous ftp and in printed form. WordNet is available in several different packages, based on computer platform.

WordNet is available via ftp, or you may use and/or ftp WordNet using a World Wide Web browser such as Mosaic or Netscape. We prefer that you ftp the WordNet system via anonymous ftp from one of the following ftp sites:

In the US: clarity.princeton.edu [128.112.144.1]

In Europe: ftp.ims.uni-stuttgart.de [141.58.127.61]

E-mail address: wordnet@princeton.edu

British National Corpus Distribution Begins

This corpus is the end-product of a unique three-year collaboration, involving Oxford University Press, Longman, Chambers-Harrap, Oxford University Computing Services, Lancaster University and the British Library, with funding from the DTI and SERC. It contains 100 million words, from over 4000 different texts carefully selected to give maximal coverage of the varieties of modern British English, both spoken and written. The corpus is automatically tagged for part of speech, using the CLAWS stochastic parser developed at UCREL, and marked up in SGML, following the TEI Guidelines for corpus encoding.

The corpus is currently available under academic licence within the European Union only. The first release, comprising three CDs and a detailed technical manual currently costs under 200 pounds.

A full installation occupies about 4 Gb of disk space, and can only be carried out on a Unix system. Later this year we hope to announce availability of the BNC Sampler: a 2 million word sample from the corpus, using an enhanced word-class tagset, manually corrected. This Sampler will be usable on standalone PC.

For full details, including ordering and licensing information, please see our web pages at http://info.ox.ac.uk/bnc or write to: British National Corpus, Oxford University Computing Services, 13 Banbury Road, Oxford OX2 6NN (Tel: +44 (1865) 273 280; Fax +44 (1865) 273 275; E-mail: natcorp@oucs.ox.ac.uk

Dutch Newspaper Corpus

The Institute for Dutch Lexicology INL offers you the possibility to consult a text corpus of over 27 million words of Dutch newspaper text, by the international computer network. In 1994, a 5 Million

Words Corpus with diversified composition has been made accessible in a similar way. With some limitations due to copyright, the output of your searches can be transfered to your own computer by e-mail. It is not allowed to transfer complete texts or substantial text parts. Most of the data has not been corrected, neither on the level of the text, nor on the level of POS and headword. POS and headword have automatically been assigned to the word forms in the electronic text by lingware developed at the INL. Additional information: Institute for Dutch Lexicology INL, P.O. Box 9515, 2300 RA Leiden. (Fax: 31 71 27 2115; E-mail: Helpdesk@Rulxho.Leidenuniv.NL)

RELATOR Language Resources Server

The RELATOR language resources server was built with the sponsorship of the Council of the European Commissioners, by the RELATOR project, for the European Language Resources Association.

The RELATOR server uses wide-area networking software to support distribution of natural language processing resources. It provides WWW, FTP, resource installation on the AFS, and email consulting services. This enables users to get an overview of speech and natural language resources very quickly, identify and compare resources of possible interest, ftp them if they are available through RELATOR and, for AFS clients, actually run pre-installed executables on the client machine for testing or development purposes.

Resources that are currently available through RELATOR include speech and text corpora, lexicons, natural language processing programs and tools, and related databases and systems. The process of acquisition and expansion is ongoing, so watch this space for further developments.

The most effective way to get an overview of natural language resources, whether available through RELATOR or not, is to browse the multilingual Web pages (http://www.XX. relator.research.ec.org, where XX stands for the two-letter country codes of the EU countries, such as de, uk, etc.)

Many resources can be obtained from RELATOR by ftp. Because the RELATOR file system is replicated across several sites, you can choose the EU site nearest you by substituting the appropriate two-letter country code for "de" (Federal Republic of Germany) in the ftp address below.

Resources can also be accessed through AFS. AFS has the additional advantage that executable files can be shared transparently across compatible hardware: this means that AFS clients can directly run preinstalled programs.

Resources available only to ELRA members, or that must be purchased or licenced before use, will not be visible at these interfaces until the necessary permission is obtained.

ftp://de.relator.research.ec.org/relator

afs: /afs/research.ec.org/projects/relator

ELRA membership information: info@relator.research.ec.org

advice on finding relevant resources: advice@relator.research.ec.org

system support: systems@relator.research.ec.org

the RELATOR network management: secretariat@relator.research.ec.org

Postal Address: RELATOR Secretariat, Centre for Cognitive Science, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, Scotland (Tel:+44 (0)31 650 4594; Fax: +44 (0)31 650 4587)

New from the Linguistic Data Consortium

Air Travel Information System

This set of discs contains a corpus of speech and natural language data collected under the auspices of the Advanced Research Projects Agency Spoken Language Systems (ARPA-SLS) technology development program. The corpus, which contains data in the Air Travel Information Services (ATIS) domain, was designed by the ARPA-SLS Multi-site Atis Data COllection Working (MADCOW) group and was collected by five sites at locations across the U.S.:

BBN Systems & Technologies, Cambridge, MA
Carnegie Mellon University, Pittsburgh, PA
MIT Laboratory for Computer Science, Boston, MA
National Institute for Standards and Technology, Gaithersburg, MD
SRI International, Menlo Park, CA

The corpora on this set of discs is part of the third phase of collection of ATIS data (ATIS3) and comprises the development test (NIST Speech Disc 17-4.2) and evaluation test material (NIST Speech Disc 17-5.1) used in the December 1994 ARPA SLS Benchmark Tests. As in the previous ATIS corpora, the speech contained in this corpus was elicited by presenting subjects with various hypothetical travel planning scenarios to solve. The resulting spontaneous spoken queries were recorded as the subjects interacted with partially or completely automated ATIS systems to solve the scenarios. Note that the ATIS3 training data is available on NIST Speech Discs 17-1.1-17-3.1.

The recorded speech has been transcribed and annotated with categorizations and canonical reference answers. All of the utterances on these discs have been recorded using a close-talking, noise-cancelling head-mounted Sennheiser microphone. For some subjects, secondary (noisier) microphone data was recorded simultaneously as well.

These discs also contain the ATIS3 46 city/52 airport relational database, a revised Principles of Interpretation, and test implementation and scoring instructions as well as other general documentation.

The ATIS3 corpus has been verified, collated, documented, and produced on CD-ROM by the National Institute of Standards and Technology (NIST) in cooperation with MADCOW and distributed by the Linguistic Data Consortium (LDC).

PhoneBook

PhoneBook is a phonetically-rich, isolated-word, telephone-speech database, created because of (1) the lack of available large-vocabulary isolated-word data, (2) anticipated continued importance of isolated-word and keyword-spotting technology to speech-recognition-based applications over the telephone, and (3) findings that continuous-speech training data is inferior to isolated-word training for isolated-word recognition.

The goal of PhoneBook is to serve as a large database of American English word utterances incorporating all phonemes in as many segmental/stress contexts as are likely to produce coarticulatory
variations, while also spanning a variety of talkers and telephone transmission characteristics. We anticipate that it will be useful in ways analogous to TIMIT/NTIMIT.

The core section of PhoneBook consists of a total of 93,667 isolated-word utterances, totaling 23 hours of speech. This breaks down to 7979 distinct words, each said by an average of

11.7 talkers, with 1358 talkers each saying up to 75 words. All data were collected in 8-bit mu-law digital form directly from a T1 telephone line. Talkers were adult native speakers of American English chosen to be demographically representative of the U.S.

Given the large set of talkers being recruited for PhoneBook database, it made sense to exploit the opportunity to collect additional utterances. We have chosen spontaneous numerical utterances, because of widespread interest in them and the need for very large numbers of talkers for research into spontaneous-speech effects. We restricted to just three spontaneous digit sequences and one money amount, as the lists for the core of PhoneBook have been designed to approach the limit of reasonable duration for a caller's session. As a result, PhoneBook contains a total of 5105 spontaneous utterances.

TRAINS Spoken Dialog Corpus

This CD-ROM contains a corpus of task-oriented spoken dialogs. These dialogs were collected as part of the TRAINS project, a project to develop a conversationally proficient planning assistant, which helps a user construct a plan to achieve some task involving the manufacturing and shipment of goods in a railroad freight system. The collection procedure was designed to make the setting as close as to human-computer interaction as possible, but was not a "wizard" scenario, where one person pretends to be a computer. Thus these dialogs provide a snapshot into an ideal human-computer interface that would be able to engage in fluent conversations.

Altogether, this corpus includes 98 dialogs, collected using 20 different tasks and 34 different speakers. This amounts to six and a half hours of speech, about 5900 speaker turns, and 55000 transcribed words.

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. The LDC e-mail address is: ldc@unagi.cis.upenn.edu.

USER SURVEYS

MT Users and Usage: Europe and the Americas

Colin Brace, Muriel Vasconcellos, and L. Chris Miller[1]

[Presented at MT Summit V, July 1995]

Introduction

Does anyone actually *use* machine translation? That's a question which has reappeared at regular intervals throughout the more than four decades in which people have tried to program computers to translate from one human language to another. The short answer is "yes," people do not as much as some might wish, but more than many believe. Success using MT has proven to be an elusive goal for many users; early adapters have not always remained faithful adherents. Recent developments, however, are radically altering this landscape.

All indications are that 1995 is shaping up to be a watershed year for MT. In the two years since Summit IV, the market for machine translation has seen a virtual revolution, and this has

greatly affected the ways in which MT is being used. The sea change for the MT market in the last two years is that prices have dropped and the platforms are such the technology is now accessible to almost anyone. MT is now also readily available through several on-line services. These changes have given people more flexibility in approaching the MT option. Because so many more modalities are now available, it can be said that in 1995 the hallmark of MT usage is *diversity*.

An Overview for 1995

For Summit V, we have endeavored to sketch a rough map of MT usage in Europe and the Americas.[2] Our overview is based in part on an official survey we conducted in the first half of 1995 with partial funding by the International Association for Machine Translation (IAMT) (see Annex 1). In this survey we attempted to gain insight into the current MT experience and to identify trends in MT use. The first IAMT-sanctioned survey of MT users was conducted in 1993, and the results were reported at Summit IV in Kobe (Vasconcellos 1993). The latest survey has helped us to identify a variety of innovative applications, and it has shown us how some of the new products that have appeared are affecting the uses to which MT is being put.

The responses this time were considerably fewer in number and less cohesive than two years ago. In 1993 they clearly pointed to certain trends, and by and large the trends correlated with the characteristics of the MT products being used. Thus, for example, in 1993 the more "mature" systems were responsible for the major share of the work being done. For the most part, the large-scale uses of general-purpose MT systems involved technical manuals and other aspects of localization, and most often the direction was from English into several other languages. The respondents found that the use of MT systems with filters that preserved graphics and format markup was saving them from 30% to 50% over their former production costs. On the other hand, the MT packages then being sold for PCs were mostly being used in casual applications, and often not for business at all. In business, they were being used to meet a variety of small-scale translation needs. In 1995, however, there is a much wider range of uses, and the trends are less consistent. The PC systems are now being used in large-scale production and are becoming important components of the business environment.

In 1993 the number of responses (38) was large enough to risk an estimate of total MT usage in the world – namely, 380 million words, or 1.2 million pages. However, a much larger sample would be needed in order to confirm this figure, which is undoubtedly only a very small share of the world's total volume of translation – perhaps less than 1%.[3] Because of the modest response this year, it is less possible than ever to estimate the percentage of translation in the world being supported by machine translation.

All in all, the 1995 survey provided us with a representative snapshot of MT use rather than a comprehensive quantitative picture. We think the response rate may reflect that MT is significantly less exotic than it was two years ago and people are less motivated to profile their operations. We also speculate that as machine translation achieves a certain stature there may be less pressure to keep records of productivity. And of course, as prices drop dramatically there is less need to justify the investment in the purchase or license of the system.

Trends

We observe four significant trends in MT today:

- An explosive growth in PC-based systems;

- An upsurge in the use of MT on-line;
- More diversified uses of MT;
- A gradual merging of machine translation proper and machine-*aided* translation.

MT on PCs

Serious machine translation is now available in affordable PC packages. In the first half of 1995 two of the MT behemoths, SYSTRAN and Intergraph, became available for the Windows environment in low-cost versions which retain all the power and capacity of their workstation or mainframe predecessors. Both these product lines, SYSTRAN Professional for Windows and Intergraph's Transcend, have the capability to handle large-scale production, and between them they cover a wide range of language combinations (see Annex 2 for details).

Globalink, in turn, which has already thoroughly penetrated the market with its PC-based line, is about to unveil a new generation of its translation software for Windows, which has all the characteristics of a full-scale transfer system and includes a user-friendly tool that allows the user to write rules for the handling of complex grammatical structures.[4] The Danish system Winger 92 is also now available in an "expert" version that allows the user to modify the existing grammar.

In another recent development, SPANAM and ENGSPAN, the systems developed by the Pan American Health Organization for internal use and the first of the mainframe systems to be ported in toto to the PC environment in 1992, are now being licensed to a limited number of public institutions.

Finally, IBM/Deutschland has teamed up with a German publisher specializing in electronic publishing to market a low-cost (DM199-DM499) English-German translation package on CD-ROM, called Personal Translator. It is derived from LMT (Logic-based Machine Translation), a system that has been under development at various sites within IBM since the mid-1980s, and it is targeted at the non-professional market. Even before the package had been released in May of this year, IBM had received over 1,000 orders for it in Germany.

Also, we should not forget that there are several PC-based systems for English-Japanese and Japanese-English: EJ Bilingual and LogoVista are being marketed in the United States for US$795 and stripped-down versions for even less.

What makes these products significant in the picture being painted here is their tremendous power coupled with their very low prices. While it had long been expected that companies with transfer-based mainframe and workstation systems would be porting their products to the desktop, most pundits doubted that the new products would have all the same capabilities or that they would be affordable to the mass market. To give an idea how wrong they were, the 1991 Ovum Report (Engelien and McBryde 1991) predicted that by the year 2000 the average MT system would be selling for US$150,000. As it turned out, the most expensive of the new heavy-duty PC systems is priced at only US$1,495. In other words, they are selling for less than 1% of the predicted cost.

Even before these new products were available, the low prices of the PC-based MT packages already on computer store shelves had been attracting users in droves, especially in the United States. The rush to reach the mass market was led by MicroTac's Language Assistant series, now owned by Globalink, which is priced at US$59 for the Windows version and sells on the street for even less. It would appear that the lower the cost, the more people are willing to try MT: all-time total sales for the Language Assistant line amounted to 400,000 units, not including upgrades or returns, at the end of 1994 when the two companies merged. (Compare with the Ovum prediction

that annual sales would reach 400 units worldwide by the year 2000 – at the US$150,000 price point).

Prices of other MT products will continue to fall as systems are increasingly bundled with other software and/or distributed at minimal cost on CD-ROMs. With over 40 PC-based MT products on the market at prices starting as low as US$39, people no longer hesitate to try out MT both in new ways and for uses that had previously been reserved for the "mature" systems. A list of PC-based MT products, current as of the date we went to press, is found in Annex 2 at the end of this report.

Yet another recent trend that is helping to boost the PC-based MT market is the availability of output in the form of synthesized speech (not to be confused with speech translation, in which the input as well as the output is in voice form). The public seems to be eager for the opportunity to hear spoken translations!

MT On-Line

Another impressive development, with potentially far greater impact, is that MT is now reaching large numbers of users in Cyberspace. In the past, raw output from SYSTRAN, ATLAS, PIVOT, and Globalink has been offered for sale on-line with varying degrees of acceptance. Now MT is being deployed to enable subscribers of on-line services to communicate with one other. In 1994 the Intergraph system was introduced in the MacCIM support area on CompuServe, and March 1995 saw inauguration of its World Community Forum, where MT enables users chat in English, French, German, or Spanish, with additional languages in the pipeline. Some 15,000 subscribers had joined this forum by the end of the first four months, which is especially significant because the connect time is chargeable over and above the basic cost of Compuserve membership. At least half of them are now regular users, and they have already posted some 26,000 messages. This experiment is reported elsewhere in detail at the present conference (Flanagan *infra*).

At the same time, both CompuServe and Globalink are currently gearing up to offer not only "raw" MT on-line but also fully postedited translations on demand. Translation customers will soon be able to submit their texts to one of these on-line services from any desktop and charge the cost to a credit card.

In another innovative use of MT on-line, the Global Schoolhouse, an offshoot of the U.S. National Science Foundation's Global SchoolNet Foundation, uses Globalink's Language Assistant series on the Internet to bring together students from 16 schools around the world. The students write stories about themselves and their families in English, French, Spanish, and Italian, which are then compiled in the form of a newsletter and exchanged.

Diversification

In 1993 it was still possible to speak of two basic categories of MT users: "industrial-strength" users who used high-end systems largely for the production of technical manuals, and casual users who used inexpensive products to extract or communicate the content of short documents. While the publication translation of technical documentation undoubtedly remains the major use of MT, lots of other interesting applications are emerging, and many of these focus on putting translation tools in the hands of non-professionals – i.e., non-translators, and are capturing the content of texts which would otherwise remain untranslated. For example, PC-based MT is now being used to sort documents by keyword in an application not unlike the military and industrial monitoring that

heretofore could only be performed by the largest of the MT systems.

This trend does not mean that skilled human translators are being displaced – in fact they are likely to have their hands more than full for the foreseeable future. Rather, it fulfills what is often referred to as the "latent demand" for translation.

Desktop Integration

With the surge in popularity of inexpensive PC products, the PC desktop is something that vendors of the well-established UNIX workstation systems can no longer ignore. SYSTRAN, LOGOS, and METAL have all been rendered local area network-aware. SYSTRAN offers a network version of its new PC-based products, while LOGOS and METAL (Max Delta is the company now handling METAL) offer Windows client software which allows PC-based users of these two Unix-based systems to submit texts remotely for translation. It is now almost universally recognized that MT has to be brought to the user; this means, in essence, bringing it to the PC-based word-processing environment of commercial packages like WordPerfect, Word, and AmiPro.

From a different perspective, machine translation and machine-*aided* translation are gradually becoming indistinguishable, by necessity, as the once monolithic high-end MT systems become more interactive and make their way to the desktop. In turn, traditional desktop translation aids such as terminology management tools, translation memory, and on-line reference works are being integrated and made MT-ready. This has led to impressive "translation workstation" packages which try to improve the overall ergonomy of translation software within the document production cycle by capitalizing on the respective strengths of automatic and manual translation.

This development has an intriguing geographical dimension: while the bulk of commercial MT development is largely a U.S. phenomenon, virtually all the major translation support packages have been developed in Europe. The result is a flurry of transatlantic activity, in which, for example, Eurolang (France) is working closely with LOGOS (USA) to render its Optimizer package LOGOS-friendly; Trados (Germany) has recently demonstrated its Translator's Workbench package harnessed to InterGraph's (USA) Transcend; and SYSTRAN (USA) has teamed up with IBM to create an environment using the latter's (Europe-based) Translation Manager. Initiatives such as Microsoft's OLE and, more MT-specific, Globalink's MTAPI, are likely to hasten this process.

Users – Europe

By now we all know that many languages are spoken in Europe and that Europe faces an awesome multilingual challenge. But language technology in general – and MT in particular – has been slow to find its way to users here. In the PC-based arena, structural differences have made the kind of success Globalink has enjoyed in the United States difficult to duplicate: European markets are smaller and more fragmented; there are fewer distribution channels; users are more conservative; and the needs of European users may not always correspond with the priorities of U.S.-based software developers.

In the high-end arena, vendors of workstation and mainframe systems have had European customers for many years, but one gets the impression that there are more *ex*-users than *current* users of MT in Europe. This impression was confirmed by two respondents to our questionnaire who identified themselves as ex-users of high-end systems. Their comments were quite revealing. "Actually, we have never really used it in the production of documents, but rather as a `toy'," concedes one ex-user, who installed a high-end system in 1988. This user cited hostility among

translators towards the system and poor support by the system's developer as contributing factors to its "failure." A representative of another company well known for its use of MT since the mid 1980s cited "inability to handle the processes involved" for joining the ranks of ex-MT users.

That being said, a small but steadily increasing volume of Europe's huge, virtually indeterminable volume of translation is done each year with the aid of translation tools of varying sophistication. As we observed above, Europe is home to four of the leading translator's workbench packages: Eurolang's Optimizer, Trados' Translation Workbench, IBM's Translation Manager, and STAR's Transit. All these packages were developed by translation and documentation practitioners, and they reflect, to varying degrees, the clear dictates of production environments. The IBM system, for example, was used for many years internally at IBM's internationalization center in Denmark before being released as a commercial product. Aimed primarily to meet the needs of teams of professional translators, these products are considerably more costly than inexpensive stand-alone packages, which are geared more for casual users. While the latter may find even the rudimentary linguistic capabilities of the inexpensive packages a godsend at given moments, translation professionals have entirely different requirements. Basic comprehension of a raw source text is the least of their problems; they may come to rely on a workbench package more for its project management and repetition handling facilities (particularly among multiple users) than for whatever linguistic smarts it may offer.

LOGOS in Europe

A couple of years ago the venerable MT developer LOGOS received an influx of German capital, and it now has a very active team based near Frankfurt which is marketing the system and supporting European users. LOGOS' customer base in Europe appears to be steadily growing, and this may be partly stimulated by their close cooperation with the French company Eurolang to offer a LOGOS/Optimizer package.

One of LOGOS' chief customers is the Language Services unit of the Swedish concern Ericsson, which uses the software to translate large technical manuals from English to Spanish, French, and German. Three of Ericsson's translators postedit all the output. In contrast to Ericsson, Osram, based in Munich, provides LOGOS to non-translators – in this case, engineers – to enable them to produce English versions of technical reports for in-house reports. Rather than postedit the output, the Osram engineers tend to tweak the input to get the translation right.

From Russia with Stylus

MT has been pursued in Russia for decades, and while the resulting systems may not enjoy the polish of their slick American counterparts, Russian MT systems are being used in Russia – and elsewhere. The St. Petersburg firm MT Project reports that more than 3,500 copies of its PC-based STYLUS have been sold. Interpol/Moscow, for example, uses STYLUS for quick translations of telexes, e-mail, and regular mail from English into Russian in the course of police work. At LONIIS (St Petersburg Telecommunications Research Institute), translators and engineers use STYLUS to process some 30,000 pages per year (English to Russian) and 2,000 pages (Russian to English) of telecommunications and computer texts.

METAL at the Technological Edge

Over the years, Sietec (a subsidiary of Siemens) has built up a small group of METAL users in

Europe. The most notable of these is the German software giant SAP, which runs large volumes of its software documentation through METAL, achieving enviable productivity gains. Sietec demonstrated a Russian-German prototype of METAL last year. METAL was recently taken over by Max Delta.

IBM Joins the Fray

We mentioned earlier that IBM/Deutschland has gotten into the PC-based MT market in Germany. An IBM spokesperson noted that the launching of this product had also stimulated interest in their more sophisticated UNIX version. On a different front, IBM is actively marketing its VoiceType dictation system in conjunction with its Translation Manager package; this combination has attracted considerable attention among the sizable number of professional translators who prefer to dictate translations.

Users – Americas

The picture in the Americas continues to be dominated by the two very different approaches to MT usage that we have mentioned already: heavy-duty use of "industrial-strength" systems, and casual use of PC-based systems. However, we do already see some of the PC systems playing a serious role in businesses and other activities. "Serious" could mean that there are multiple users, that the system is being used on a regular basis or even full-time, or that MT is making a difference in the way the user's work is accomplished. The advent this year of full-scale systems at affordable prices is certain to create many intermediate points along the casual-to-heavy-duty spectrum.

With these systems, the true test of MT's acceptability to users in the Americas is about to be faced.

The PC Phenomenon

PC-based MT packages are selling like hotcakes in the Americas. The explosion that we described above has been felt more strongly in the United States than anywhere else in the world. They are being sold to a lesser extent in Canada, Mexico, and Latin America. It is no coincidence that these products have flourished in a highly organized mass market that is avid for PC software; a number of them might never have seen the light of day in other circumstances. In fact, it would not be unreasonable to say that the popularity of the Language Assistant series – which has led the trend in terms of volume sales – has been more of a marketing phenomenon than a linguistic one.

The evidence of users' acceptance and satisfaction with PC-based MT products is still largely anecdotal. Increasingly they are being used for occasional business correspondence. Many people start to use the inexpensive packages to make themselves understood in a language that is foreign to them, which then sets up an exchange in which MT is also used for understanding the message that is received in return.

In the 1993 report we heard about an American priest using MT to prepare a sermon in French, and about tourists carrying around their PCs on their travels. There was also a man who was finally able to communicate with his relatives in Italy. This year we learn that an American admirer of Italian sourdough bread has translated the original recipes using Language Assistant and is now able to make the same bread that he had "eaten and loved in Italy." Students of MT evaluation will remember that cooking recipes are the performance test par excellence of a translation.

Large-Scale Users

In the meantime, the high-end, heavy-duty MT systems continue to be used to process large volumes of translation. The break point that appears to justify using MT is an output per language combination of at least 1,000,000 words (4,000 pages) of translation per year. In the majority of cases, MT is being used for technical documentation, whether produced by the manufacturer or by an independent translation agency. Although these users seem to be satisfied with the role that MT plays in their operations, their number does not appear to have increased very much in the last two years.

LOGOS has a few new users in the Americas as well as in Europe. One of them reports "high volume output of decent quality if the text is sufficiently simple in structure" but complains that it is "difficult to convince and `evangelize' skeptics who judge MT only by output quality in comparison to human translations without taking the whole process into consideration."

The caveats about simplicity of style apply mostly to producing translations of publication quality. There are also several long-term users of MT in the Americas that rely on MT to convey the information content of large volumes of unconstrained input over a large range of subject areas. The most venerable example is the U.S. Air Force, which after 25 years is still so satisfied with SYSTRAN that it continues to underwrite the development of new source language modules and cover the cost of conversion to up-to-date platforms.

Another satisfied information-gatherer switched in 1991 from Weidner's old MICROCAT to Intergraph's DP Translator. This company also uses MT for publication-quality translation. They report a 30% increase in productivity for the latter type of text and gains of 50% and more in the case of information-only translations. They note that, despite using MT for nearly a decade, there is a "continuing requirement for substantial time commitment to dictionary updating."

One of the other long-term users is the Pan American Health Organization, which has had MT for 15 years. PAHO developed its own systems from Spanish to English (SPANAM) and English to Spanish (ENGSPAN) and ported them to the PC in 1992. The translations cover a wide range of subjects and the input texts are not controlled or pre-edited in any way. Into English the annual volume is 1,400,000 words (86% of the service's total volume in that language direction) and into Spanish, 1,100,000 (70% of total volume). PAHO's success may well be due to the fully supportive environment, from specially tailored word-processing macros to daily updating of the MT dictionaries.

Conclusion

Thanks to MT, a small but growing percentage of professional translation activities in several major languages is being automated. MT is also making it possible for texts to be translated for informal or non-publication use which would otherwise remain untranslated. So we can safely conclude that MT is serving a useful albeit modest function in the area of multilingual communication.

But MT is not for everyone. In the vast majority of cases, highly skilled human translators continue to outperform even the best of today's systems when publishable texts are required. Reasonable productivity gains can only be achieved in a very small number of domains in which the syntax is naturally constrained – technical documentation is the best example. Tremendous amounts of work on all these systems have yet to be done to make them adaptable to additional domains. And while MT vendors have been making great strides in getting MT closer to users' desktops, there are still gaps in the levels of integration, particularly with regard to terminology management and MT

lexicons. For now, it seems likely that a certain segment of the products on the market will continue to serve casual users, and that the "professional" and "expert" versions will remain the tools of translators and other specialists. However, eventually the two types of systems are likely to merge into a single entity.

So what does the future of MT look like? Well, it runs on a PC and it costs less than five hundred dollars...

NOTES

1. E-mail: Brace, colinb@ibm.net; Vasconcellos, 71024.123@compuserve.com; Miller, 70303.314@compuserve.com.
2. The Asia-Pacific region was included in 1993 survey, but the Summit V program divided up the task by region, and as a result the survey and present report have been limited to Europe and the Americas.
3. If there are 175,000 working translators in the world (Van Slype 1983), and they produce an average of 1,000 words a day for 200 days in the year (a low figure for a full-time professional and a high one for the casual translator), the total volume in the world in a given year would be 35 billion words, or 140 million pages).
4. The new Globalink product line, code-named "Barcelona" at this point, offers full-scale transfer MT from English to Spanish, French, German, and Italian and will be ready for beta-testing in the third quarter of 1995.

REFERENCES

Engelien, B., and R. McBryde. 1991. *Natural Language Markets: Commercial Strategies*. London: OVUM, Ltd., 1991.

Van Slype, Georges. 1983. *Better Translation for Better Communication: A Survey of the Translation Market, Present and Future.* Oxford: Pergamon Press.

Vasconcellos, Muriel. 1993. "The Present State of Machine Translation Usage Technology, or: How Do I Use Thee? Let Me Count the Ways." In *MT Summit IV: Proceedings* (Kobe, 20-22 July 1993), pp. 35-45.

Annex 1

1995 International Survey of MT Users

The 1995 International Survey of MT Users was an official survey conducted and compiled with partial funding from the International Association for Machine Translation (IAMT). The first IAMT-sanctioned survey of MT users was conducted in 1993, and the results were reported at Summit IV in Kobe (Vasconcellos 1993).

For the report to the present Summit, the authors updated the 1993 survey using basically a two-pronged approach: first, obtaining current information from the users who had responded two years earlier, and second, ferreting out new users. With support from IAMT, the responses collected in 1993 were entered into a database, which will now be maintained on a regular basis.

As a next step, the respondents in Europe and the Americas were presented with their entries from two years ago and asked to bring them up to date.

In the effort to find new users, several strategies were enlisted. As before, the commercial vendors were contacted and asked for client names. At the same time, appeals were posted on CompuServe and on various places on the Internet. In addition, continued networking yielded further leads, and, finally, the authors followed up on examples of MT usage that they had been gathering in their own files over the intervening two years.

In order to have comparable data between systems and over time, we repeated the same questions that had been asked before. They were kept to a minimum so that they would not be too formidable:

Survey Questions

System used? Since when?
Language combinations (from ___ into ___)?
Hardware platform? Since when?
Form of input (e.g., disk, downloaded files, OCR, manual keying)?
Purpose of translation?
Type of documents translated: discourse genre (e.g., "technical manuals"), subject matter?
Output per year (number of words) percentage of total translation volume?
Dictionary size (number of entries) for each language combination?
Description of personnel who use it (e.g., contract translators, etc.)? How many?
Type and amount of pre-editing done?
Type and amount of postediting done?
System for incorporating feedback from end-consumers?
Advantages, disadvantages of MT?
News flash: Latest developments? Novel uses of MT?
Plans for the future?

Annex 2

Commercially Available PC-Based MT Products

Compiled by *L. Chris Miller*

COMPANY	AppTek, Inc.
CONTACT INFO	1420 Beverly Road, Suite120, McLean, VA 22101 (703) 821-5000 Fax (703) 734-5703 APPTEK@CLARK.NET
PRODUCTS	AppTek EAMT for UNIX or SUN
LANGUAGES	English into Arabic
IN DEVELOPMENT	Arabic into English, Translator Workstation for Windows

---

COMPANY Bilingual Software
CONTACT INFO P.O. Box 292700 Davie, FL 33329-2700
Sales (800) 232-8228 Information (305) 434-2721 Fax (305)434-5604

PRODUCTS Translate INSTANT SPANISH v2.0 DOS $169, Pronto Spanish $69 CDROM only
LANGUAGES English into Spanish
IN DEVELOPMENT Windows version expected in June 1995

COMPANY EJ Bilingual Inc. - Nippon Texa Co., Ltd./Kimihira and Taylor Associates, Inc.
CONTACT INFO 2483 Torrance Blvd., Suite #1, Torrance, CA 90501
(310) 320-8139 Fax (310) 320-3228

PRODUCTS EZ JapaneseWriter v.2.0 Windows $795, v 2.1 DOS $595 Basic version $299
LANGUAGES English into Japanese

COMPANY Globalink, Inc.
CONTACT INFO 9302 Lee Highway, 12th Floor, Fairfax, VA 22031
(800) 255-5660 (703) 273-5600 Fax (703) 273-3866
Compuserve Vendor J Forum or info@Globalink.com

PRODUCTS Spanish, French, German, and Italian Assistant v.5.1 DOS, v.1.0 Windows, v.1.0 MAC $59.00, localized Windows versions $79
Spanish, French, German, and Italian Assistant Deluxe Windows CD-ROM $59
LANGUAGES Bidirectional Spanish/English, French/English, German/English, Italian/English
PRODUCTS Power Translator DOS v.2.0 $89
LANGUAGES Bidirectional English/Spanish, English/French, English/German
PRODUCTS Power Translator DOS v.2.0 $89
LANGUAGES English into Chinese
PRODUCTS Power Translator Windows v.2.0, MAC v.2.0 $249
LANGUAGES Bidirectional English/Spanish, English/French, English/German
PRODUCTS Power Translator Deluxe for Windows CDROM edition $299
LANGUAGES Bidirectional English/Spanish, English/French, English/German
PRODUCTS Power Translator Professional DOS v.3.0 $299
LANGUAGES Bidirectional English/Spanish, English/French, English/German, English/Russian
PRODUCTS Power Translator Professional Windows or Mac $595, OS/2 $199
LANGUAGES Bidirectional English/Spanish, English/French, English/German
PRODUCTS Dictionaries for Power Translator Professional DOS $59,
Windows, Mac or OS/2 $89
IN DEVELOPMENT Italian, Portuguese

COMPANY INFO:Partner A/S

CONTACT INFO Arnold Nielsens Boulevard 68, DK 2650 Hvidovre, Copenhagen, Denmark
+45 3677-1810 fax +45 3677 1905

PRODUCTS Winger 92
LANGUAGES Danish-English, English-Danish, Spanish-English, English-Spanish

---

COMPANY Intergraph Corporation
CONTACT INFO 1 Intergraph Way - GD3001
Huntsville, AL 35894-001
(800) 222-9242 (205) 730-9832 Brazil 55 11 887-5300 Mexico 52 5 207-5262

PRODUCTS Transcend for Windows $495 per language direction $795 Bidirectional
LANGUAGES English into Spanish, Spanish into English, English into French,
French into English
IN DEVELOPMENT German into English, English into German, English into Italian,
English into Portuguese (shipping June 1995)

COMPANY Language Engineering Corporation
CONTACT INFO 385 Concord Avenue, Belmont, MA 02178
(800) 458-7267 (617) 489-4000 fax (617) 489-3850

PRODUCTS LogoVista E to J v.2.1 Japanese Windows or MAC or Power MAC $1995
LogoVista E to J Personal v.1.5 $795
LANGUAGES English into Japanese
PRODUCTS 21 subject-specific dictionaries $495/$995
IN DEVELOPMENT English to Japanese E-mail translation service.
Japanese into English expected w/in 1 year

COMPANY Linguistic Products
CONTACT INFO P.O. Box 8263, The Woodlands, TX 77387
(713) 298-2565 fax (713) 298-1911

PRODUCTS PC-Translator DOS v3.3 $985 per language pair $1585 Bidirectional
LANGUAGES English into Danish, Dutch, German, French, Italian, Norwegian, Portuguese,
Spanish, and Swedish. Danish, Dutch, French, German, Italian, Norwegian,
Portuguese, Spanish, Swedish into English

COMPANY Road Scholar
CONTACT INFO 2603 Augusta, Suite 1000, Houston, TX 77057
Sales (800) 336-5989 Support (713)266-7623 Fax (713) 266-4525

PRODUCTS Spanish Scholar v.2.0 Windows $49.95
LANGUAGES Bidirectional Spanish/English

COMPANY SMART Communications, Inc.
CONTACT INFO 885 Third Avenue, 29th floor, New York, NY 10022
(212) 486-1894 Fax (212) 826-9775 jsmart@interport.net

PRODUCTS Smart Translator V3.1 Site licenses $25,000+ (custom client/server solutions)
LANGUAGES Bilingual English into Castillian or Latin American Spanish, European or Canadian French, German, Italian, Creole
IN DEVELOPMENT English into Mandarin Chinese

COMPANY Socatra
CONTACT INFO 5500 Royalmount Ave., #320, Town of Mount-Royal, Quebec, Canada H4P 1H7
(514) 735-7079 Fax (514) 735-9697

PRODUCTS XLT v.3.0 UNIX $5,000 initial subscription, additional language pairs $2,500
Microchip access rate from 2 to 3 cents a word. Annual renewal $1,000.
LANGUAGES English into French, French into English
IN DEVELOPMENT English into Spanish or Italian, Spanish or Italian into English
Windows expected within 1 year

COMPANY Softkey International, Inc.
CONTACT INFO 450 Franklin Road, Suite 100, Marietta, GA 30067
Sales (800) 227-5609 Support (404) 428-0008 Fax (404) 427-1150

PRODUCTS Key Translator for Windows v.1.0 Disks or CDROM $39.95
LANGUAGES Bidirectionall English/Spanish

COMPANY SYSTRAN Software, Inc.
CONTACT INFO 7855 Fay Avenue, Suite 300, P.O. Box 1926, La Jolla, CA 92037
(619) 459-6700 Fax (619) 459-8487 info@systranmt.com

PRODUCTS Systran Professional for Windows standalone version $1495 per language pair; network version $2495 5 user, 10 user $3995, 20 users $7495
LANGUAGES English into Spanish, French, German, Italian. Spanish, French, German, Japanese into English.

PRODUCT Systran Professional for Windows standalone version $995
LANGUAGES English into Portuguese
SUPPORT 60 days unlimited technical support via toll-free hot line, Internet, fax, and BBS

Yearly maintenance agreement $225 Standalone, $375 5 users, $600 10 users, $1125, 20 users

IN DEVELOPMENT Korean, Chinese, Russian, etc.

COMPANY Westcliff Software
CONTACT INFO 343 Soquel Ave. #207, Santa Cruz, CA 95062
Sales through CLR (800) 900-8803

PRODUCTS DosAmigos DOS 5.0 $199.00 Spanish Amigo Windows v.2.0 $99.00
LANGUAGES Bilingual Spanish/English

COMPANY Character Language Resources
CONTACT INFO 2130 Sawtelle Blvd., #304-A, Los Angeles, CA 90025-6250
(800) 900-8803 (310) 996-2300 Fax (310) 996-2303

Distributes comprehensive assortment of multilingual software, including many Machine Translation software products. This is a good resource for MT products that are difficult to find, such as:

PRODUCT ESP Plus for Windows $349 Requires Chinese Windows
LANGUAGES Bidirectional Chinese/English

PRODUCT Targumatic DOS $119 per language pair $199 Both pairs
LANGUAGES Hebrew into English, English into Hebrew

PRODUCTS JE Bank DOS, Windows, Mac $795
LANGUAGES Japanese into English

PRODUCTS PC-Transer ej/je Windows (Win-V or Japanese Windows) or Mac (with System 7.1 and Japanese language kit) $1995 per language pair.
LANGUAGES English into Japanese, Japanese into English

PRODUCTS PC Transer Junior (with Twinbridge, Win-V, or Japanese Windows) $129
LANGUAGES English into Japanese

PRODUCT LTGold DOS $169
LANGUAGES Russian into English

PRODUCT LexiTrans DOS $385
LANGUAGES Bidirectional Russian/English

PRODUCT STYLUS for Windows $585
LANGUAGES Bidirectional English/Russian, German/Russian

IN DEVELOPMENT Bidirectional French/Russian

The information on this list was believed to be accurate at the time of writing. The inclusion of a product should not be taken as an endorsement. Comments are welcome and should be addressed to L. Chris Miller via mail at 2020 Penn. Ave., N.W., Washington, DC 20006, fax (703) 780-1822, or by electronic mail at 70303.314@Compuserve.com.

SURVEY REPORT: Korya Eiwa

100-dollar commodity has expanded the market

Sadao Hoshino
(AAMT Secretary-General)

[From AAMT Journal no.11, June 1995 (unedited, except spelling)]

Introduction

In November, 1994, a MT software commodity appeared in the market which destructed the market price so drastically far beyond the standard of the conventional translation softwares. It aroused a big sensation in the MT industry. Named "Korya Eiwa", or "It's Nice! English-Japanese" with the maker's retail price 9,800 Japanese yen. Then, the Association Secretariat felt it necessary to investigate the influence which that commodity exerted. Cooperation was requested to the developing vendor of "Korya Eiwa!", and the user survey was executed by inserting questionnaire sheet in each set of the system and distributed to the retailers.

Method and subject of investigation

The investigation adopted the questionnaire method. The questionnaire form was packed up together into the commodity of the April job-lot. Those answered sheet were sent back as fast as in May. The answers of 110 from among them were totaled according to the publication date of the AAMT Journal. However the totals seem to be a little insufficient, the user trend can be analyzed effectively with those samples.

Originally, I wanted to survey the users for the period six months following the purchase of MT systems, when the problem of use can be understood more clearly. It should be noted that the questionnaire for this survey has collected answers of the unskilled user occupying the majority of the samples. Thus, there are a lot of "NA"s and "Uncertain" in the answer. It is a pity that because the customer list had not been disclosed to us by the vendor, the follow-up check was not able to be done.

Survey result

The investigation result of the utilization realities of the personal-computer MT software suggests the danger that only the numerical value will be emphasized and referred. The point is, that our survey does not necessarily indicate the general or overall status of how all the personal-computer MT softwares are actually utilized by which user in what market. This survey report, to the last, merely remains as the investigation of how actually a single commodity is applied by its users. Therefore, it is necessary to understand the target/goal as well as the characteristic of this commodity enough, as much as it is necessary to analyze the numerical value in adequacy.

A detailed explanation of this commodity was printed in the preceding issue. Please refer to AAMT journal No.10, March 1995 [reprinted in this issue]

Outline of commodity's target

- to offer Windows version. Operativeness has improved.
- Only the minimum function is loaded, since the majority of the users dislike/hesitate the editing work.
- The translation engine adopted that of a real sect with high-speed high-accuracy.
- The lexicon was upgraded to hold 67,000 headwords.

Characteristic

- The clipboard function refers to the lexicon.
- Translation equivalent/original language correspondence. Translation equivalent substitution.
- Looking up of inflection into a dictionary. Vague expression retrieval function. Job History function.

The user answered immediately after the purchase of the MT software. The user did not select the commodity based on the characteristic of the commodity at all. They do not use each and all functions supplied. However, the standard of the above-mentioned characteristic judges the total numerical value.

User's trend

The user demanded the improvement of the MT softwares ready on the market, and to institute those problems to the software developer/vendors, the Association has fed back the investigation result to them. The analysis of each item was omitted here, and the analysis is entrusted to each reader.

Outline of the PC MT users

Firstly, the outline of the user trend is described. Of all, the user has extended to all regions in Japan. So far, the translation system for WorkStations (WS) had been sold by the house-to-house, or house-calls system, and the main target of the sales activity was those enterprises in the larger cities. Meanwhile, the WS MT systems have been quite expensive. They required thorough instruction for fully utilizing them, and after-sales service were necessary. They were mainly used for larger amount of translation processing of the manualsand handbooks.

On the contrary, the MT softwares for the personal-computers (PCs) are far cheaper, and the PC retailers chiefly sells them at the shop front. Therefore, the user has expanded from the larger cities to the provinces (refer to Q.7).

User's age group has changed

The price of MT software for PC is very handy. The teenage users accounted for 7% of the total questionee [i.e. respondents]. The one's in his/her twenties was 13%: Those aged above fifty only consisted 4% of the total, suggesting that the higher aged is inexperienced/unfamiliar with PC yet. While a part of local public entity aggressively enhance PC education for citizens, the change in the surrounding environment will have the aged citizens to reconsider their PC influence in the future.

No translation experts

Next, 92% of the purchaser of MT PC is an enthusiastic, inexperienced with translation (Q.5). Considering the internationalization, they want to read/absorb/make use of the flooding English information aggressively. Student user is 17% of the total. There is a student in his/her

eleventh grade, or he/she is 16 or 17 years old. The teachers occupied 4% of the total. Attention must be paid to those housewives as well as the unemployed (refer to the "Job" section).

Private and office users

As the answer of the translation purpose overlaps, the translation demands to enjoy personal hobby reaches 32%. For the translation original, magazine 14%, novel art and literature 8%, letter 10%, and newspaper 10% all suggests the private use of PC MT.

Among the answers, some bought PC MT to aid the English education of the child in the household. Therefore, although a person tries to answer to all questionnaire items, there are those items they could not answer, as they are not the exact users of the PC MT.

The MT is used for the study purpose as well. The students answer frequently that they apply PC MT for "Thesis decipherment". While "Input through the telecommunication network" reaches 20% of the input mode, it can be guessed that the use of overseas data base is also actively persuaded (7%).

In addition, the translation of the catalogs (20%) and the manuals/handbooks (36%) stands out, too (Q.12). Is it because of the increase of imported merchandises and popularization of importing overseas goods/services privately?

There are two religionist users, too. As more and more overseas information is supplied extensively, the wish to collect data is naturally heightened, thus lifting the demand for the MT software. For reference, though the source is rather old, only 2,875 titles, or 4.7%, were translated into Japanese, among the world's 61,531 translated book titles in 1986. Religious books in that were a mere 118 titles.

Influence of the sale of $100 PC MT software

The sale of this commodity largely influenced the commodity sales in this industry, with other companies experienced decline in shipment for two to three months. However, this new product has expanded the width of the user segment recently. Therefore, its effect in total even promoted the sales of the higher-ranked more expensive MT softwares in the industry. The industry evaluated the contribution of this new product extremely highly.

Afterwards, a number of low-priced softwares were commercialized following "Korya Eiwa!". The merchandise line of the PC MT softwares was also sub-divided and enhanced with the professional, standard and popular versions. Now, the varied selection of commodity is supplied for the users and potential users of MT on PCs, which will certainly contribute to a large expansion of the market quite effectively.

Questionnaire results (Unit:Person/matter)

Q.1. What kind of PC do you apply the MT software on? (Multiple answer accepted.)

Macintosh type: 4
MS-Windows (NEC's PC98 series): 61
MS-Windows (DOS-V version): 55
Others: 4

Q.2. What MT software do you use? (Multiple answer accepted.)

Korya Eiwa (English-Japanese MT): 110.
Others: 10.

Q.3. For what purpose do you usually use PC for your business? (Multiple answer accepted.)

Documentation: 79
Table calculation: 54

Translation: 37
The data preservation: 32
Others: 38
NA: 2.

Q.4. Please evaluate your English fluency.
Fluent. Can read thesis: 4
Good. Can speak: 21
Not fluent at all: 82
NA: 3.

Q.5. Translation experience (years)
Less than one year: 0
Between three to four years: 2
More than five years: 2
No experience: 101
NA: 5

Q.6. Experience of MT software/system previously.
Yes. Less than six months: 6
Six months or more: 1
No: 103.

Q.7. Shop/vendor where translation software was bought.
Co-op shop: 1
Personal-computer retailer: 84
OA (office automation) supplier: 14.
Software retailer: 7
Others: 3
NA: 1.

Q.8. Purpose of MT application. (Multiple answer accepted.)
For one's business: 12.
For private use: 71.
For private hobby: 35.
For study/learning: 1.
Others = for trial before buying more expensive software: 1.

Q.9. What motivated the purchase and use of MT. (Multiple answer)
The translation demand for business increased: 4.
The translation demand has heightened: 16.
To make the best out of one's hobby: 41.
To raise the job efficiency: 45.
Others = for trial before buying more expensive software: 1.
Others = without specification: 2.

Q.10. Degree of translation accuracy expected. (Multiple answer)
All sentences must be translated completely: 7.
Accurate enough to understand the outline of the text: 56.
Accurate enough to understand the summary: 34.
To obtain the translation equivalent for the domain specific terminology: 10.

As an aid for studying English: 20.
NA: 2.

Q.11. Domain of the text to be translated. (Multiple answer accepted)
Transportation: 1.
Telecommunication: 2.
Plant and factory: 1.
Construction: 1.
Computer Science: 24.
Semiconductor: 2.
Machinery: 3.
Shipbuilding: 1.
Electricity: 8.
Electronic industry: 5.
Quality control: 1.
Law and patent matters: 2.
Pharmacology: 1.
Automotive products: 1.
Metal: 1.
Chemical industry: 1.
Science studies: 1.
Biology/biochemistry: 2.
Physics: 2.
Medicine: 2.
Economics: 1.
Astronomy: 1.
Mathematics: 1.
Musicology: 1.
History: 1.
Education: 1.
English studies: 1.
Theology: 1.
Religious studies: 1.
The unemployed 2.
NA: 37.

Q.12. Category of translation originals. (Multiple answer accepted)
Legal documents: 1.
Theses: 17.
Patent documents: 3.
Manual/handbooks: 58.
Instructions: 40.
Contracts: 9.
Reports: 11.
Order documents: 3.
Catalogs: 22.

Estimate/quotations: 2.
Guide books: 15.
Standardization: 7.
Newspapers: 11.
Magazines: 15.
Novel and art and literature 9.
Reports 17.
Letters: 11.
Data base contents: 7.
Electronic mails: 24.
NA: 5.

Q.13. Amount of translation.

Number of pages per translation work converted into A4, or letter sized, form:

1 to 3 pages: 55.	4 to 5 pages: 12.
6 to 10 pages: 7.	11 to 20 pages: 4.
21 to 50 pages: 2.	51 to 100 pages: 1.
No idea: 4.	NA: 22.

The average number of sheets translated per month.

1 to 10 pages: 23.	11 to 20 pages: 29.
21 to 50 pages: 12.	51 to 100 pages: 2.
101 to 200 pages: 2.	More than 1,200 pages: 1.
No idea: 4.	NA: 37.

Q.14. Medium of translation original. (Multiple answer accepted)

Most are hand-written: 11.
Printed: 65.
Digitized = Recorded on magnetic medium: 28.
Obtained through telecommunication: 26.
Various media: 2.
NA: 2.

Q.15. Method of inputting translation original. (Multiple answer accepted)

With OCR devices: 18.
Through telecommunication line: 22.
FD input: 28.
Keyboard input: 63.
Others: 6.
NA: 5.

Q.16. Situation of pre-editing. (Multiple answer accepted)

Longer sentence is divided into shorter sentences: 4.
Replenish the missing subject of the sentence: 4.
Replenish the missing object of the sentence: 1.
The sentence structure is corrected: 2.
Useless words deleted: 2.
The part of speech is specified: 1.
The proper translation equivalent is selected: 8.

Spelling is checked: 13.
The inserted phrase is specified: 0.
The original text is rewritten: 8.
Do not pre-edit at all: 51.
NA: 28.

Q.17. Number of domain specific lexicons possessed.
Technological engineering: 1.
Computer science: 1 (50,000 headwords)
Electronics: 1.
Arabian language: 1 (50,000 headwords)
All fields: 1 (60,000 headwords)
NA: 91.
Uncertain: 2.
Do not possess: 12.

Q.18. The number of headwords in the user dictionaries. The number of headwords entered in the parenthesis.)
Automotive products: 1. (----)
Mathematics: 1 (1,000)
Computer science: 1 (300)
CPU: 1 (----)
All fields: 1 (500)
No user dictionary: 104.
Uncertain of the entry: 1.

Q.19. If the user dictionary/lexicon is shared with others or not.
Shared among the members of the same job/class group: 3.
Shared among friends on the telecommunication network: 3.
Not share with others: 77.
NA: 27.

Q.20. Processing of unknown word left in the translated text. (Multiple answer accepted)
Leave as it is: 53.
Replace with its katakana notation of its pronunciation: 14.
Replace with a personal translation equivalent: 23.
Others: 3.
NA: 24.

Q.21. Functions supplied with "Korya Eiwa" not used so often. (Multiple answer accepted)
Lexicon: 1.
Other printing font application: 1.
Printing out: 1.
Recitation: 2.
HELP function: 1.
NA: 105.

Q.22. Time-consuming process. (Multiple answer accepted)
OCR input: 1.
OCR input mistake check: 1.

Pre-editing: 1.
To correct the grammatical mistake of the original: 1.
To translate the domain specific terms properly: 1.
Translation of idioms: 1.
User dictionary entry: 6.
Selection of proper translation equivalent: 1.
Post-translation check: 2.
To translate those plural words written in row in the original text: 2.
Pre-/Post prcessing of signs and logos: 1.
To pre-edit and shorten lengthy sentences: 1.
Translation is done by each sentence, thus requires post-editing on paragraph basis: 1.
Do not find anything inconvenient: 1.
Not sure: 1.
NA: 95.

Q.23. The reuse/application of translation sentence at other occasions/purposes.
Often reuses/apply: 39.
Hardly uses them again: 51.
Others: 5.
NA: 15.

Q.24. Evaluation of translation software including "Korya Eiwa". (Number of answers in the parenthesis.)
[ITEMS: A=Good or satisfied; B=As expected C=Dissatisfied; D=Not sure; E=NA; F=Uncertain]

Operativeness:
A=(30). B=(64). C=(7). D=(3). E=(5). F=(1).

Translation quality.
A=(4). B=(46). C=(49). D=(4). E=(7). F=(0).

Translation speed.
A=(31). B=(53). C=(15). D=(5). E=(6). F=(0).

Applicability to various processors.
A=(15). B=(49). C=(4). D=(36). E=(5). F=(1).

Handiness of pre- and post-editing.
A=(10). B=(49). C=(26). D=(19). E=(5). F=(1).

Cost effectiveness to function.
A=(67). B=(28). C=(4). D=(5). E=(6). F=(0).

Q.25. Functions desired to be equipped (Multiple answer accepted).
Study function: 1.
Automatic pre-editing: 1.
OCR function: 1.
Real-time translation while telecommunicating: 1.
MS-DOS version: 1.
Spelling checker: 1.
The direct translation of OCR input: 1.
Translation of conversational style sentence: 1.
Recitation: 1.

AI functions: 2.
Japanese-English translation: 1.
All sentence batch translation: 1.
Phonetics input: 1.
Enhancing better domain specific lexicon: 6.
Study aiding system for junior high school English: 1.
Presentation of another interpretation function: 1.
Deletions of all English translation sentences: 1.
Partial re-translation: 1.
Connection/data interchangeability with word processing softwares: 1.
Automatic entry of corrected translation equivalent to the lexicon: 1.
The compound word is specified on the screen: 1.
Automatic document polish/rewriting function: 1.
The translated text file preservation: 1.
Lexicon registration of any words with any part of speech: 1.
Translation with transfer analysis: 1.
File preservation of both English (original text) and-Japanese (translated text): 1.
Input of data other than text files: 1.
Uncertain: 1.
Not needing any: 1.
NA: 61.

Q.26. The purchase of higher-ranking MT software.
Will consider: 48.
The current state is satisfactory: 32.
Not consider: 19.
NA: 11.

Questionee's age group analysis.
11-19 years old: 8.
21-29: 14.
31-39: 36.
41-49: 33.
51-59: 2.
Older than 60: 2.
NA: 15.

Job/occupation of the questionee.
White collared office worker: 33.
Independent enterprise/self employed: 4.
Housewife/home maker: 1.
Student: 19.
Religionist: 2.
Medical doctor: 2.
School teacher: 4.
Civil servant: 4.
Manufacturing industry employee: 1.

Corporate officer, private firm: 1.
Professional translator: 1.
Unemployed: 2.
Others: 11
NA: 25.

READERS' FORUM

Machine Translation and the Office Automation

Key-Sun Choi

[From AAMT Journal no.11, June 1995 (unedited)]

The office automation (OA) is for efficient interpersonal communication. This communication has two aspects: control and information. The control has two processes: workflow control and cooperation control. On the other hand, the information automation leads to two problems: information sharing and information extraction. Terms of "sharing" and "control" assume the standardization for communication.

The workflow means a sequence of document processing. Here, a document is a representation of information. Initialization of workflow is to search a right path. One of search methods can come from the information classification. Each process of workflow is human investigation of document or its improvement. A human involved in each step has different knowledge and different experience. Office workers have different roles on, for example, bookkeeping, information extraction, decision-making, etc. Each of them requires a different level of knowledge: for example, foreign language ability, terminology knowledge, and expert knowledge. What can MT contribute to in this situation? What can MT help them, each of who has different knowledge level and different viewpoints or purpose? How is MT adapt to such workflow environment?

MT should be integrated with information retrieval. This approach contributes to the right application of workflow to foreign documents. MT should also be integrated with knowledge base for supporting the terminology and expert knowledge. However, a reader of document cannot interpret completely, whereas the writer of the document can interpret it completely.

We can see one solution, that is, document interchange format (DIF). DIF contains the basic terminology and writers' analyzed knowledge. This means that each document contains the standardized interchange format for right translation. Incoming DIF includes the knowledge to interpret it. The initial phase of workflow interprets an incoming document according to the local knowledge base plus the attached DIF. Why should each document contain such complex information? The authors of document can best analyze the right meaning of the document. DIF has several levels according to the use level of recipients. This means that the different use requires different contents of DIF.

The different user role and their knowledge level require a different MT strategy. DIF contributes to such flexible MT strategy.

Key-Sun Choi, Center for Artificial Intelligence Research, Korea Advanced Institute of Science and Technology, Taejon 305-701 Republic of Korea. Email: kschoi@cair.kaist.ac.kr

Human-quality Fluency in Machine Translation!

Bruce Wydner

If your company does or wishes to do a considerable volume of communications over the Spanish-English language barrier and you have been among those investigating developments in Machine Translation to see if they may offer you any significant cost or time savings or any promise of an "Interlingual Internet Service" – that permits ALL of your people with NO linguistic training to communicate back and forth over this language barrier with FULL UNDERSTANDING through Machine Translation with "human-quality Fluency" – we have what you want.

[My Associates and I have prepared this Service for virtually all of the world's written languages into English and back, but at present we are only taking orders to deliver it for the Spanish-English language barrier.]

We do this through a constant, extensive Statistical Analysis of a customer's usage of this Service coupled with an apparently still exclusive understanding of the Principles behind the developments in Natural Language Processing that were inaugurated when my invention of cost-effective Machine Translation, the Weidner Multilingual Word Processor, was introduced on the market in 1978.

We are not able to achieve Human-quality Fluency in the Machine Translation of your company's specific usage without this Statistical Analysis. However, if you would care to send us any selection of the Machine Translation of Spanish into English messages done by MT software we will show you or your experts how we utilize these Principles to eliminate many of the cost and time consuming shortcomings of these software products which result in their Fluency being so far below the level of Human Quality.

Reply to Bruce Wydner (102436.1606@compuserve.com)

PUBLICATIONS ANNOUNCED AND RECEIVED

International Journal of Corpus Linguistics

The *International Journal of Corpus Linguistics* (IJCL) is a new journal that will be launched in 1996, under the editorship of Wolfgang Teubert and published by John Benjamins Publishing Company (Amsterdam/Philadelphia).

IJCL will present a wide range of views on the role of corpus linguistics in language research, lexicography and in natural language processing (NLP).

IJCL seeks to publish research that views language as a social phenomenon that can be investigated empirically on the basis of authentic spoken and written texts. Corpus linguistics specifies corpus design in respect to research interests, provides computational methods of extracting linguistic knowledge, and conceives tools to validate the accuracy of linguistic description. It views meaning as language usage to be studied in actual corpus citations. Corpus linguistics provides help to compare languages and to analyse language change. IJCL will discuss aspects such as the above.

IJCL focuses on corpus-based monolingual and multilingual lexicography. Corpus

linguistics provides a methodology to deal with multi-word units, collocations, set phrases, etc. Comparable and parallel corpora can supply context-sensitive translation equivalents better than standard dictionaries. IJCL covers corpus-based dictionaries for human users as well as lexicons for NLP applications.

IJCL aims to conciliate the expectations of language industry with the goals of academic linguistics research. Corpora are the basic resources in language engineering. It is the linguistic knowledge extracted from corpora that determines the performance of any NLP application. IJCL is a forum to exchange and share experience, expertise, visions and information on resources and tools.

IJCL will, besides a substantial information section, also feature book reviews, abstracts of relevant articles in leading journals and in important conference proceedings, reports on focal language and language technology centres, both in the academic and the industrial field.

For more information : John Benjamins Publishing Co., Attn. IJCL, P.O. Box 75577, 1070 AN AMSTERDAM, The Netherlands (Tel: +31.20.6762325; Fax: +31.20.6739773; E-mail: kees.vaes@benjamins.nl)

Communications of Chinese and Oriental Information Processing Society

This journal is intended to be published 4 issues per volume. Its scope includes all aspects related to Computer Processing of Chinese & Oriental Languages, e.g., computer input and output of characters, typesetting and design of characters, coding and compression of data, voice input and output, analysis, recognition and synthesis of speech, man-computer communications, language processing and text understanding, representation of knowledge and inferencing, computational linguistics, machine translation, software and design of Chinese language computers, database management and systems, information retrieval, text handling, question answering, applications of theories, methods and techniques.

Subscription rates: S$30 per year for 2 issues for Singapore and Malaysia; S$50 for all other destinations. Further information: COLIPS, c/o DISCS, National University of Singapore, Kent Ridge, Singapore 0511 (E-mail: luakt@iscs.nus.sg)

IEEE Computer
Theme Issue on Interactive Natural Language Processing

Computer has planned to devote the July 1996 issue to Interactive Natural Language Processing. Manuscripts reporting survey, original research, design and development, and applications of Interactive Natural Language Processing are sought immediately in the following areas:

+ Speech Understanding and Generation Platforms
+ Natural Language Interfaces and User Interface Management Systems
+ Dialog/Discourse Management and Story Understanding Environments
+ Interactive Machine Translation systems (and Translator's Workbenches)
+ Intelligent Writing Agents

The particular focus of this theme issue is the special considerations in theory and practice of "real-time" processing of natural language as opposed to "batch-mode" processing. How do the theoretical and processing assumptions differ when the input stream is "live", thus increasing the emphasis on

response time and throughput? What are the trade-offs? What are the mechanisms (symbolic, statistical, connectionist, hybrid)? Finally, what are the strategies that enable interactive natural language processing to either be effected or finessed?

Information: Bill Z. Manaris, Computer Science Department, University of Southwestern Louisiana, Lafayette, LA 70504-1771 (Tel: (318)482-6638; Fax: (318)482-5791; Email: manaris@usl.edu)

Machine Translation: two special issues

Special Issue on Lexical Choice in Text Generation and Machine Translation

The journal of *Machine Translation* is inviting submissions for a Special Issue on Lexical Choice in Text Generation and Machine Translation. This is the second special issue of the MT Journal that explicitly addresses Text Generation. The first issue (edited by Richard Kittredge, Montreal) was devoted to research in Text Generation that has been applied to Machine Translation or to multilingual language generation. This Issue is devoted to a specific problem -- that of lexical choice – which arises in both Text Generation and Machine Translation regardless of the underlying theoretical model and the application area of the research.

Although lexical choice is of great importance to both Text Generation and Machine Translation, it has very often been ignored as semantic constructions have tended to be associated directly with lexical units. Only recently has lexical choice started to become one of the major areas of research in Generation, and it still remains a significant unresolved issue in Machine Translation. The problem of lexical choice is particularly difficult because it is inseparably intertwined with the problems of syntactic realization, discourse generation, knowledge representation, lexicon organization, etc. This means that solutions to lexical choice require that other tasks in generation and machine translation have already been solved to some degree.

This Issue addresses all critical topics in the problem of lexical choice in Text Generation and Machine Translation. One of the basic topics is the placement of the lexical choice process in the

generation and translation process, and, subsequently, the range of phenomena that the lexical choice process has to deal with. Another basic topic is the acquisition of constraints for the guidance of the lexical choice process. Until now, lexical choice processes have only rarely "consciously" decided upon the selection of one out of several possible lexicalizations – although it is generally accepted that decisions are to be made in accordance with (target) language-constraints and culture-specific constraints. An additional basic topic is the most suitable organization of semantic and lexical resources for lexical choice.

High quality research papers are invited on these and other topics, including but not limited to:

- Finding appropriate lexical equivalents in Machine Translation.
- The role of lexical choice in Multilingual Generation.
- Acquisition of lexical resources relevant to lexical choice.
- Organization of semantic resources and the lexicon for lexical choice.
- Design and implementation of lexical choice processes.

- Interdependency between lexical choice and the other tasks in generation and machine translation.

- Phenomena to be addressed by lexical choice.

- Contextual criteria for lexical choice.

FORMAT: Articles may be submitted in hard-copy or electronic (plain ASCII or .ps) format to the guest editor. If submitting hard-copy, four copies of the paper are required. The length of the papers should be approximately 20-30 pages (12-point font).

DEADLINE: Submissions are due on September, 15 1995

GUEST EDITOR: Leo Wanner, Computer Science Department, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (Tel: +1/519-888-4567 5344; Fax: +1/519-885-1208; E-mail: lwanner@after.logos.uwaterloo.ca)

Special Issue on New Tools for Human Translators

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

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

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

Some promising concepts have been discussed in recent literature, such as:

* automated terminology extraction and management
* translation memories (including bilingual concordancing systems)
* translation checkers
* translation dictation systems

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

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

DEADLINE: February 29, 1996.

FORMAT: Please send hardcopy or electronic file (plain text or PostScript) to either guest editor: Kenneth Ward Church, AT&T Bell Labs office 2B-421, 600 Mountain Av., Murray Hill, NJ 07974, USA (E-mail: kwc@research.att.com); Pierre Isabelle, CITI, 1575 Chomedey Blvd., Laval, Quebec, Canada H7V 2X2 (E-mail: isabelle@citi.doc.ca)

The 1996 Language Engineering Directory

Paul M.Hearn

The 1996 *Language Engineering Directory* contains information on hundreds of organisations and commercially available products in the language technology field. The study from which it results is part-financed by DG-XIII of the Commission of the European Communities in Luxembourg.

Application areas covered in the Directory 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 other categories.

The 1996 edition will be available towards the end of this year. If you would like to kept in touch concerning the exact publication schedule, please return your full name, address and telephone number by e-mail or snail-mail and I will see you are placed on the mailing list.

Alternatively, if your institution markets directly or indirectly any products in the above fields, I would be happy to receive information on these products, for possible inclusion in the Directory this year.

Address: Paul M. Hearn (Editor, The 1996 Language Engineering Directory), Postbox 2.572, 28080 Madrid, Spain (Tel/fax: 34.1.519.9198; E-mail: 100575.2522@compuserve.com

Translation and Text Technology

[From TRANSST no.22 (January 1995)]

The new series *Translation and Text Technology* aims at disseminating a variety of new research and engineering approaches to knowledge-based translation and text technology. This technology is characterized by a particular concern for meaning and knowledge representation and partly overlaps the fields of artificial intelligence and machine learning. It clearly exceeds the boundaries of linguistic morphology or mere string handling, addressing problems related to semantic proximity, recognition of contextual similarity and knowledge-based disambiguation. Though it may reveal important theoretical insights, the work in this area is primarily motivated by the construction of practical systems and instruments, such as translators, talkwriters, adaptive help systems, advanced word processors, high-quality speech generators, etc. Therefore, special attention will be given to the use of large resources for semi-automatic knowledge acquisition (text corpora, dictionaries, term banks) and the application of case-based or corpus-based techniques.

The series will be edited by Claude Bédard, Brian Harris, John Laffling, Klaus Obermeier, Jun-ichi Tsujii and Toon Witkam. It will be published by Mouton de Gruyter.

Ovum publishes Globalisation report

Globalisation: creating new markets with translation technology, [by] Rose Lockwood, Jean Leston [and] Laurent Lachal. London: Ovum, 1995 (ISBN: 1-898972 25 7)

This is the first of two reports on the MT market to be published by Ovum Ltd. this year. [The second entitled "Language technology" has not yet been received by MTNI.] In its 412 pages

this publication is the most thorough and detailed survey of the market for products and services supporting language translation and multilingual globalisation. The authors report on the factors which are influencing the formation of this market and make forecasts about potential growth up to the year 2000 worldwide.

Globalisation is defined in the report as the planning of products for an international market, the adaptation of products and services to local conditions, the development and manufacture of products in the most appropriate locations, the maximisation of market opportunities, and the use of new technology to bind together globally distributed organisations. Many of the barriers to globalisation are being removed with changes in trade agreements, multi-national markets (e.g. the European Union), and new telecommunications; but the most significant barrier is not technical but cultural: the sensitivity to local needs and expectations, above all to the local language; and at its heart is, therefore, translation.

The report considers four categories of 'translation products': language-reference products (such as dictionaries and language-checking tools), language resources (such as terminology-handling systems and terminology databases), machine-translation systems, and translation-memory systems. It covers globalisation services, localisation and translation vendors, translation bureaux, and freelance translators. And it considers four types of user groups: software publishers, manufacturing companies (producing technical documentation), information industries and service companies, and the public sector.

The authors estimate the present (1995) market for translation products as $200 million world-wide, and predict that it will grow by nearly 50% per annum to $1.5 billion in the year 2000. The globalisation services market is larger: at present ca. $1.9 billion, predicted to grow to $6.3 billion by the end of the decade. The report provides predictions for each of the product and service categories: e.g. language reference and resource tools are forecasts to grow from $60 million (1995) to $655 million (2000), and MT products from $112 million (1995) to $573 million (2000). The most rapid increases are predicted in the globalisation services for software (from $564 million to $2390 million) and for technical documentation (from $660 million to $2020 million); by contrast, services in the public sector will grow very slowly, from the current $277 million to $354 million in 2000.

There are interesting implications for the vendors of MT products in these predictions: language translation aids will overtake MT systems as such in their share of the market, and the main use of systems will continue to be for software and technical documentation. It should be stressed, of course, that the Ovum report does not investigate some markets for MT products outside the business and manufacturing community, e.g. the purchase and use of cheaper PC-based systems by non-translators, nor does it consider the potential markets for automatic spoken language translation, but those it does cover are of central and crucial importance to the whole MT community. Above all, MT vendors should read closely the recommendations about what users would like from their tools, what features they would like to be improved and introduced into future products and services.

Apart from its detailed examination of the globalisation processes, the services and the products available, the report includes valuable profiles of leading users (e.g. Aerospatiale, Caterpillar, CompuServe, Ericsson, Ford, SAP, Sybase), service vendors (e.g. AT&T Business Translations, Berlitz Translation Services, Cap Volmac Lingware Services, International Translation & Publishing Ltd., McQueen, Mendez Translations SA, Rank Xerox, Donnelly (Stream

International), Trantex), and product vendors (e.g. Eurolang, Fujitsu, Globalink, IBM, INSO (InfoSoft), Logos, Sietec, Smart Communications, Trados).

It is intended to publish more substantial reviews of both Ovum reports in the next issue of MT News International. For information about purchase contact Ovum Ltd., 1 Mortimer Street, London W1N 7RH (Tel: +44 171 255 2670; Fax: +44 171 255 1995).

Ward, Nigel: **A connectionist language generator** (Norwood, NJ: Ablex, 1994)

This book is a description of a small-scale but ambitious project to build a generator capable of dealing with a wide range of knowledge sources. At present, it is argued, most generators are designed to accept inputs relatively poor in information: feature structures, lists of propositions, logical forms, realization specifications. But if generators are to produce genuinely idiomatic output appropriate for specific contexts, they will have to deal with a wide range of information-rich information: genre, register, roles, attitudes, intentions, level of familiarity, settings, prior probabilities of events, and encyclopedic information about objects, concepts, events, and persons. They will have to be sensitive to nuances of meaning which human speakers deal with effortlessly. Nigel Ward describes a connectionist model (FIG) which in principle can perform with information-rich sources in real time and incrementally. The context of the model is translation from Japanese into English, and for this reason it is of interest to MT researchers. It is readily admitted to be a 'toy' which addresses only a small fraction of the task, but it is the first attempt to design a connectionist generator which integrates interdependent lexical, syntactic, and semantic sources, which processes simultaneously and in parallel different sources of knowledge, which does not build structures, and which incorporates feedback mechanisms during output. This thought-provoking book should be read by anyone with an interest in pursuing connectionist approaches to problems of natural language generation.

PUBLICATIONS RECEIVED

Journals

AAMT Journal *no.9 (December 1994).* Contents: Japanese picture scrolls and multi media communication (Manabu Watanabe). -- What really needs for interlingua-based MT (Huang Changning). -- Research report Indonesian side (Darmawan Sukmadjaja). -- Another decade of Thai language language processing research (Virach Sornlertlamvanich). -- MMT in a teleworking environment (Mohd. Zaini Yunus). -- Automatic interpretation system and speech recognition technology (Shinsuke Sakai). -- Machine translation system for European languages and English: Power Translator from Iris International Co. -- LogoVista E to J Personal released from LogoVista Corporation. -- NTT Communication Science Laboratories, Machine Translation Research Group (Satoru Ikehara). -- Machine translation revolutionalizes the industrial translation (Takeshi Narumi)

AAMT Journal *no.10 (March 1994).* Contents: Machine translation is available (Masataka Yamamoto). -- Translating technical materials with PC MT software - my personal challenge and suggestions (Tanehiro Tatsuta). -- NHK Science and Technical Research Laboratories (Terumasa

Ehara). -- An overview of text readers (Yoshiaki Kurosawa). -- Research results and future directions of EDR (Toshio Yokoi and Hiroshi Suematsu). -- PC-Transer series for E/J and J/E translation by Nova, Inc. --Japanese-English machine translation software: TransLand for Windows, Brother Industries, Ltd. -- A Windows-based translation tool "Korya Eiwa!", Catena Corp. -- Denjikai for Windows: an English-Japanese and Japanese-English dictionary software by Fujitsu Co., Ltd. -- WD-01SW, an English text translation support software for word processor WD-MF01 - Sharp Corporation. -- PENSEE for Windows: a Japanese-English and English-Japanese translation support software, Oki Electric Industry Co.Ltd. -- j.London e/j for English to Japanese translation, Japanese Windows 3.1 version by Kondensha Co.Ltd. -- MT software/systems for PCs. -- Study of machine translation in Singapore (Dong Zhen Dong).

AAMT Journal *no.11, June 1995* [Japanese only]. English diskette contains: MT User survey: $100 MT for PCs () [see p. ?]. -- MT and office automation (Key-Sun Choi) [see p. ?] -- Technology for improving sentences: a short guide for MT techniques, part 14 (Toshikazu Fukushima). -- The fifth AAMT General Members' Assembly [see p. ?]. -- Trans-supporter EJ: an E-J translation support software by Sanyo Information Business Co. [see p. ?]. -- HICOM-MT, a Japanese-Korean MT system by Hitachi Information Network Ltd. [see p. ?]. -- Announcement - future events: NLPRS'95 to be held in Seoul, Korea; LISA Forum - Asia. -- AAMT Office moved.

Computational Linguistics, *vol.21, no.1 (March 1995)*; *vol.21, no.2 (June 1995)*

Elsnews, *vol.4 no.3 (May 1995)*. Contents: ELRA now receiving members. -- Meeting to discuss integration of speech and NL held in UK [at Durham Univ.] (Russell Colliingham). -- FRANCIL: language engineering, en français (Joseph Mariani). -- THe Dutch national R&D programme "Language and speech technology" (Lou Boves and Alice Dijkstra). -- Update on HCM spoken dialogue and discourse project. -- Workshop on discourse and dialogue prosody held in Stuttgart (Ernst Buchberger)

Elsnews *vol.4 no.4 (August 1995)*. Contents: Industrial placement service to be implemented soon. -- "He that tholes, overcomes", report on the Third European Summer School on Language and Speech Communication (Kay Berkling et al.) -- Training in NL and speech: integrated individuals or integrated teams? (Yvonne van Holsteijn). -- ACSYS: making speech technology accessible in the marketplace (Khalid Choukri).

INL Infoterm Newsletter *76 (June 1995)*

Language Industry Monitor *no.24 (November-December 1994)*. Contents: A new era for Systran (Muriel Vasconcellos and L.Chris Miller). -- Ingénia-Langage Naturel: targetting French grammar (Andrew Joscelyne). -- An enthusiastic MT user: LogoVista E to J gets a workout at one small Japanese translation shop (David J.Littleboy). -- Japanese MT: not a trivial pursuit (David J.Littleboy). -- Sietec listens to Metal users (Klaus Schubert).

Language Industry Monitor *no.25 (January-February 1995)*. Contents: PolyDoc's docucentric universe. -- Lightbulbs and Logos. -- CompuServe likes online MT. -- Intergraph's Transend. --

OCR and linguistics. --Fulcrum does the 'net. -- Merriam-Webster on CDROM. -- Yet more Japanese translation aids.

Language International, *vol.7 no.3 (June 1995)*. Contents include: Globalink developments. -- Intergraph offers Transcend.

Language International, *vol.7 no.4 (August 1995)*. Contents include: PARS, MT system from Russia. -- First Italics conference (Michael Anobile). -- Finnish to English MT.

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

LISA Forum Newsletter, *vol.4 no.2 (June 1995)*. Contents: The second LISA survey: whose agenda is it anyway? (Rose Lockwood). -- Culture & convention in multimedia interface design (Steven Forth). -- The Eastern European localization business: personal observations (Arturo Quintero Arellanes). -- Sistemi Informativi Translation Center (Chiara Orlandini). -- TechTra/CCN - the principle of 'direct access' to know-how (Lotte Neiffer)

vol.4 no.3 (August 1995). Contents: High-growth markets for localisation (Rose Lockwood). -- Translating software: what it means and what is costs for small cultures and large cultures (Dai Griffiths, Stepehn Heppell, Richard Millwood and Greta Mladenova). -- Small and medium-sized enterprise software support for Bulgaria and Lithuania (Todorka Damianova and Richard Millwood). -- PRACTEAST and TACIS (Norbert Kalfon). -- The death of the amateur translator skills in the modern world (Deborah Fry). -- Symantec Houdini: how Symantec addresses some of its help localisation issues (John Rowley). -- Ovum evaluates: translation technology products (Jane Mason). -- Copernicus and the ALEP system (Iain Urquhart). -- Translating technical materials with PC MT software: my personal challenge and suggestions (Tanehiro Tatsuta)

Literary and Linguistic Computing *vol.10 no.2 (1995)*

Natural Language Engineering *vol.1 part 1, March 1995*. Contents: Technical terminology: some linguistic properties and an algorithm for identification in text (John S.Justeson and Slava M.Katz). -- Natural language interfaces to databases: an introduction (I.Androutsopoulos, G.D.Ritchie and P.Thanisch). -- Natural language processing in an operational clinical information system (C.Friedman, G.Hripcsak, W.DuMouchel, S.B.Johnson and P.D.Clayton).

Terminology, *vol.2 no.1 (1995)*.

Terminology Standardization and Harmonization: newsletter of ISO/TC 37, *vol.7 no.2 (June 1995)*

The Translator, *vol.1 no.2 (1995)*.

TRANSST: an international newsletter of translation studies. *No.22 (January 1995)*.

Tribune des Industrie de la Langue et de l'Information Électronique, *no.17-19 special issue*

(1995). Contents include: M.L.A.P ou le lever du rideau. -- Vers la société de l'information: l'appel d'offres 1994-1998 (Dominique Perrin and Virginie Boutin). -- France: politique de la traduction, politique de la langue (Olivier Lesage). -- Canada: une politique de la traduction au service de la cohésion sociale (André Abbou). -- Tous les chemins mènent à la TAO (André Abbou). -- Une sémantique nouvelle pour la traduction automatique: les classes d'objets (Gaston Gross). -- La TAO des dernières aux prochaines années... (J.M.Lange). -- Aleth Trad, outil et service (José Vega). -- Interagir pour traduire: la TAO personnelle pour rédacteur monolingue (Hervé Blanchon). -- La théorie du procédé constructif (André Abbou). -- Du marché global de la traduction à celui de la traduction scientifique et technique.. et à celui de la traduction avec outils (André Abbou). -- Les sentier actuels de la stratégie industrielle et de la stratégie marketing (André Abbou). -- Une offre en mutations successives... (André Abbou). -- Evaluation des systèmes de traitement de l'information: d'une évaluation à l'autre (André Abbou, Pascale Morey and Virginie Boutin). -- La société de l'information mondiale: perspective ou mirage? (André Abbou). -- Traduction et autoroutes d'information: un modèle pour pays volontaristes, même en temps de crise [Japan] (André Abbou). -- La recherche en interaction homme-machine en Ukraine (Françoise Néel and D.Teil). -- La langue arabe face aux industries de la langue (Christian Fuhr)

Books

Alberto, Paulo; Bennett, Paul (eds.): **Lexical issues in machine translation**. Luxembourg: European Commission, 1994. (Studies in Machine Translation and Natural Language Processing, Vol.8) 138pp. ISSN: 1017-6568. Contents: Introduction (Paulo Alberto and Paul Bennett). -- Eireterm, Eurotra's terminological database: history and development (Donncha Ó Cróinín and George Talbot). -- The computational interpretation of Germanic noun-noun compounds: an overview of possibilities and hypotheses (Lieve de Wachter). -- Automatic translation of English compounds: problems and prospects (Kerry Maxwell). -- Relational adjectives: their characteristics and correspondences (Marta Carulla). -- Lexical semantics and the problem of multiple reference in some polysemic and non-polysemic nouns (Olga Alejandro and Flora Ramírez). -- Introducing HORATIO (Archibald Michiels). -- Feeding LDOCE entries into HORATIO (Archibald Michiels). -- How far do printed dictionaries and MT-lexicons share information? (Anna Braasch)

Toury, Gideon: **Descriptive translation studies and beyond.** Amsterdam/Philadelphia: John Benjamins, 1995. viii,311 pp. (Benjamins Translation Library, vol.4) ISBN: 90-272-1606-1

Ward, Nigel: **A connectionist language generator.** Norwood, NJ: Ablex Publishing Corp., 1994. xi,298pp. ISBN: 1-56750-038-2. [see also p. ?]

Whitelock, Peter and Kilby, Kieran: **Linguistic and computational techniques in machine translation system design.** 2nd edition. London: UCL Press, 1995. xii,208 pp. (Studies in Computational Linguistics) ISBN: 1-85728-216-7

Conference proceedings

Clas, André and Bouillon, Pierrette (eds.) **TA-TAO: recherches de pointe et applications**

immédiates. Troisièmes Journées scientifiques du réseau thématique "Lexicologie, Terminologie, Traduction" Montréal, 30 septembre, 1er et 2 octobre 1993. Montréal (Québec): AUPELF/UREF, 1994. xviii,521pp. ISBN: 2-909611-09-4 [For a report of the conference see MTNI#7:1-2]

Proceedings of the Third Workshop on Very Large Corpora, 30 June 1995, Massachusetts Institute of Technology, Cambridge, Mass. Edited by David Yarowsky and Kenneth Church. 198pp.

CSNLP 1995: 4th International Conference on the Cognitive Science of Natural Language Processing, Dublin City University, 4-7 July 1995. Editor: A.I.C.Monaghan. var.pp.

Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI 95. July 5-7, 1995, Centre for Computational Linguistics, Katholieke Universiteit Leuven. 372pp.(in 2 vols.)

MT Summit V Proceedings, Luxembourg, July 10-13, 1995. [216pp.]

International Conference: Recent Advances in Natural Language Processing. Proceedings, Hotel Orpheus, Tzigov Chark, Bulgaria, 14-16 September 1995. 355pp.

Reports

Guide des produits et services d'ingénierie linguistique accessibles en France. Sous la direction d'André Abbou. Paris: Ofil, 1995. 109pp. ISBN: 2-906036-07-4 (Tribune des Industries de la Langue et de l'Information Électronique, no.16)

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 also p. ?]

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