MT News International

Newsletter of the International Association for Machine Translation

ISSN 0965-5476 Issue no.10, January 1995

IN THIS ISSUE:

Editor-in-Chief:
John Hutchins, The Library, University of East Anglia, Norwich NR4 7TJ, United Kingdom
Fax: +44 (1603) 259490; Email: J.Hutchins@uea.ac.uk; or: 100113.1257@compuserve.com
Regional editors:
AMTA: 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

AMTA Holds its First Conference

Kevin Knight
(Information Sciences Institute)

The Association for Machine Translation in the Americas held its highly successful First Conference (AMTA-94) from 5-8 October at the Columbia Inn in Columbia City, Maryland. Over 170 people registered, with a balanced representation of academic researchers and

commercial practitioners. This unique mix was instrumental in achieving the goal laid out in the subtitle of the conference, "Technology Partnerships for Crossing the Language Barrier."

The highlights of the conference were the technical presentations, panels, invited talks, tutorials, and demonstrations. Credit goes to Muriel Vasconcellos, conference chair; Eduard Hovy (USC/ISI) and Joseph Pentheroudakis (Microsoft), program co-chairs; and Bill Fry (AMTA), exhibit coordinator, for setting up a ceaseless menu of interesting ideas and MT systems. Credit also goes to Deborah Becker and Jane Zorrilla, who organized and ran the registration and information desks during the conference.

Jaime Carbonell (CMU) gave the keynote address, describing CMU partnerships that led to successful MT systems and kept research lively. Maghi King (ISSCO) described the evolution of MT partnerships in Europe, ending with the Verbmobil project and new calls from the European Community for MT development. John Hutchins (Univ. East Anglia) gave a third invited talk about the past and future of MT, including a comparison of MT approaches and funding cycles in different regions. He noted that desire for surveillance, official multi-language policy, and national competitiveness were three very different driving forces behind MT around the world. Hutchins also mentioned newly declassified Russian archives identifying the KGB as a major MT funder.

Researchers from all over the world were represented in the technical program, with topics that included: translatability, knowledge-based MT, speech translation, sublanguages, statistics, resources, and future visions. The final day of technical talks included two papers on
Japanese-English MT, two on Korean-English MT, and two on Chinese-English bilingual corpora.

Several commercial vendors and research teams demonstrated their MT systems, including: Apptek, ATR, Carnegie Group, Citac, CMU, CompuServe, Globalink, Language Systems, Logos, Language Engineering Corporation, Pangloss, PAHO, Sietec, and Systran. (Globalink, Logos, and Systran were also partial sponsors of the conference.) A new addition to the field was d'Armond Speers' Klingon Language Analyzer, which performed a challenging task as well as splitting the conference registrants into the haves and have-nots of popular culture (where do they speak Klingon?). Some of the demonstrations ran throughout the conference, while others ran in thirty-minute time slots, parallel to the technical talks, with time for questions and answers. While the technical talks dealt significantly with East Asian languages, the demonstrations tended toward European language pairs.

On the day before the technical program, there were four three-hour tutorials: Choosing an MT System (Karen Spalink), Intellectual Property Rights and MT (Leighton Chong), MT in the 1990s (Muriel Vasconcellos), and Interlingual MT (Sergei Nirenburg, David Farwell, Boyan Onyshkevych).

Five panels set the stage for lively discussion about various MT-related topics. The first was the provocatively titled "Is MT Research Doing Any Good?" Ken Church (AT&T Bell Labs) advocated closer ties between the MT community and professional translators. Bonnie Dorr (Univ. Maryland) pointed out that traditional divisions like academia/industry and research/applications do not reflect the reality of work in MT. In another panel ("Evaluation of MT"), ARPA's George Doddington made the case for evaluating core technology, in the form of fully automatic MT quality. Others on the panel stressed features like ease of use and extensibility. The "Economics of MT" panel discussed MT cost. Denis Gachot of Systran presented an ideal case of 30 to 40 percent cost savings over human translation. There was some discussion of the pros and cons of offshore human translation services, bearing on issues like up-front investment and reliability. Several speakers pointed

out that the economics of MT had much in common with the economics of human translation – many problems are problems for both.
A fourth panel covered MT in operational settings, and a fifth dealt with the future of MT. Sensing the end of the conference, panel members and audience engaged in spirited discussions of two topics: the size of the translation market, and MT versus translation aids. Pierre Isabelle (CITI) described MT's share of the translation market as 3 to 4 percent; Ken Church followed up with an exhortation to work on translation aids rather than MT. Ed Hovy likened the translation market to the market for automobiles in the early days, claiming that cost-effective improvements in MT quality would create a near-infinite demand for translations. As for the future, Jaime Carbonell suggested that MT requires a long-haul approach, in which researchers must double as developers (and triple as maintainers), so that steady progress can be made. He also encouraged research groups to spend a fraction of their time on high-risk/high-payoff ideas, such as semi-automatic language learning by instruction.

In addition to the schedule of presentations, there were several opportunities for conference participants to relax, socialize, and forge new partnerships or reinforce existing ones. Social events included a welcoming reception, a spectacular dinner cruise in Baltimore Harbor on Thursday, and a banquet on Friday night.

The next AMTA conference is scheduled for 1996, and all indications are that it will continue to be held every two years. During non-conference years, specialized workshops may be held by one or more of AMTA's special-interest groups, which are: MT interlinguas, MT lexicons, personal computer MT, MT evaluation, and MT standards.

The Proceedings of the conference have been published, and are available from the AMTA (see elsewhere in this issue for an order form).

International Workshop on "Machine Translation & Translation Theory" 14-16 September 1994, Hildesheim

Susanne Heizmann

The 2nd International Workshop on "Machine Translation and Translation Theory" was held between September 14th and 16th 1994 at Hildesheim University. The event, brought into being in 1992 by the EAMT, was this year co-sponsored by the EAMT and the German Ministry for Research and Technology, and hosted by Christa Hauenschild from Hildesheim University. Since the Ministry was represented by researchers from the VERBMOBIL speech translation project, several of the papers presented covered topics from this new field of research.

Following the 1992 approach, the Workshop was intended to stimulate the exchange of views between researchers from the two fields of Machine Translation and Translation Theory. Experts from Europe and the USA presented results of their own work and discussed possible interconnections with respect to each other. It was particularly this exchange of views that the participants judged as being especially fruitful. On the whole, the Workshop met with a positive response from all sides, and the participants voiced the opinion that it should soon be repeated.

The papers presented were the following: Margaret King (Geneva): Evaluating translation. – Susanne Petzolt, Birte Prahl (Hildesheim): Translation problems and translation strategies in human and machine translation: Empirical studies. – Susanne Jekat-Rommel, Birte Schmitz (Hamburg/Berlin): Translation equivalents in automatic dialogue interpreting. – Jan Amtrup (Hamburg): Perspectives for incremental MT. – Monika Doherty (Berlin): Aspects of informational density. – Christiane Nord (Heidelberg/Hildesheim): The

importance of functional markers in (human) translation. – Hans Hönig (Germersheim): Using text mappings in teaching consecutive interpreting. – Peter E. Pause (Konstanz): Interlingual strategies in translation. – Louis des Tombe (Utrecht): Compensation. – Susann LuperFoy (Washington): Discourse representation and processing for voice-to-voice machine translation: VERBMOBIL and the interpreting telephone. – Rita Nübel (Saarbrücken): Knowledge bases for MT. – Barbara Moser-Mercer (Geneva): Process models in simultaneous interpretation. – Hans Krings (Hildesheim): The human interface in MT: Empirical research on post-editing. – Birgit Apfelbaum, Cecilia Wadensjoe (Hildesheim/Linkoeping): How does a VERBMOBIL device affect conversation? Discourse analysis and machine-supported bilingual discourse. These and other papers on the subject of co-operation between MT and TT will be published next year.

In the lively discussions between papers and in the evenings, it became clear that an exchange between researchers from the two fields is not only desirable but can also be highly productive. Among the questions discussed were the following:

- Can process models from translation theory and discourse analysis be introduced into MT software?

- Who are the MT users of the future? or: What should MT systems and MT outputs for non-translators look like?

- Will hybrid models allowing for both microstructural (e.g. phonetic, syntactic) and macrostructural (e.g. situational) information to be processed during translation be a solution for the problems in MT as well as TT?

These and many other questions could, of course, not be answered during three days. However, the participants agreed that the discussion must be continued between and outside workshops, and that interdisciplinary co-operation is essential for those who wish to adapt to the MT market demands of the future.

Information on the planned publication and on other aspects concerning the workshop can be obtained from : Susanne Heizmann, Universität Hildesheim, Computerlinguistik, Marienburger Platz 22, D-31141 Hildesheim, Germany. (Tel.: +49 5121/883-329; Fax: +49 5121/83644; E-mail: suse@cl.uni-hildesheim.de)

Fourth International Summer School "Contemporary topics in Computational linguistics" (with emphasis on Machine Translation) 1-6 September 1994, Tuzlata, Bulgaria

Ruslan Mitkov

Following the good traditions so far, the summer school program offered a variety of interesting courses for the participants coming from 15 different countries. The summer school was opened by an 'Introduction to MT' given by H. Somers. Next, M. King presented an introduction to 'Evaluation in machine translation'. Other courses related to MT included 'Latest developments in machine translation' (S. Helmreich, S. Nirenburg), 'Machine-aided translation' (Z. Yusoff) and 'Anaphora resolution in machine translation' (R. Mitkov).

Except for the MT topics, the summer school program offered a range of various Computational Linguistics courses such as 'Introduction to semantic syntax' (P. Seuren), 'The meanings in the language, in the world, in our minds' (M. Zock), 'Discourse structure and reference resolution' (R. Delmonte), 'Automatic abstraction' (B. Tsou), 'Models for for time, tense and aspect in natural languages' (M.Zock) and 'How to synthesize words: the problem of multilingual word-form generation' (D. Cristea).

We envisage a publication of selected courses from the Summer schools held from 1992 to 1994 as a special collection entitled "Contemporary topics in Computational Linguistics".

Translating and the Computer 16
10-11 November 1994, London

John Hutchins

The sixteenth in the annual conferences organised by Aslib in London was held in the splendid 19th century headquarters of the Institution of Civil Engineers, close by Westminster Abbey and the Houses of Parliament. Some two hundred participants from Europe and some further afield came to learn about current developments in the translation technologies.

The proceedings were opened by two presentations of developments in computational linguistics and machine translation. Harry Bunt of the University of Tilburg gave "An overview of the current state of Natural Language Processing", and John Hutchins (University of East Anglia) described the changes which have taken place within the last five years in research on machine translation. The second session began with Sophia Ananiadou (UMIST) on developments in terminological research and management which are becoming increasingly important for translators. She was followed by a description of the current Logos MT system by John Hatley, and by a paper from David Canter (read in his absence by Emma Wagner) on the implications and potential impact of Sweden's membership of the European Union on the translation profession in that country.

After the lunch break, we heard from Helen McCready (IBM) about the IBM TranslationManager and the exciting newly launched VoiceType product enabling translators (and any other users of computers) to enter text by dictation. The system analyses and converts spoken words into text, and has an active vocabulary of 32,000 words. It needs to be trained to the user's voice and the speaker must pause after each word (i.e. no continuous speech), but an amazing accuracy of 95% is claimed at a dictation speed of up to 100 words a minute. Not surprisingly, the demonstration next day was packed.

The last two presentations on the first day were by Stelios Piperidis (Institute for Language and Speech Processing, Athens), who described the EU-funded research on TRANSLEARN, a corpus-based translation memory tool applied to legal texts in English, French, Portuguese and Greek, and by Lorna Balkan (University of Essex), who outlined the EU joint research project for developing test suites for evaluating MT systems.

The second day began with practical sessions and demonstrations. Chris Pyne (Sietec) described how an MT system can be successfully integrated into the workflow of an organisation. Peter Ball (an independent consultant) gave valuable advice on how to select suitable and cost-effective computer hardware and software. The demonstrations were given by Sietec and by Logos of their MT systems and translation tools, and (as already mentioned) by IBM of its TranslationManager and its new VoiceType product.

After lunch, Dorothy Senez (Commission of the European Communities) described the recent rapid growth in the use of the Systran system at the Commission primarily as a result of its promotion among officials for 'rough drafts' of short-lived documents and thanks to the provision of links to the Celex multilingual database of Community legislation and to the Eurodicautom terminology databank. She was followed by Geert Adriaens (Siemens Nixdorf and University of Leuven) describing another EU project, the LRE project SECC for developing a writing tool for non-native writers of English technical documents, based on a controlled language and with facilities for style correction. The computational framework is

the Metal MT system; the partners are Siemens Nixdorf, Cap Gemini, Alcatel Bell, the University of Leuven and Sietec; and the subject domain is that of telecommunications.

The concluding speakers were Minako O'Hagan (Victoria University of Wellington), who enthusiastically predicted a new world for translators in the 'cyberspace' of global communications, and Colin Brace (Language Industry Monitor), who summed up the current technological environment of the translation community with its multiplicity of tools from the fully automatic MT system to the humble word processor, and looked forward to a future of increasing integration and compatibility of systems.

Thus concluded another successful, enjoyable and well organised Aslib conference, with participants looking forward expectantly to next year's event. The proceedings are available from: Aslib, Information House, 20-24 Old Street, London EC1V 9AP, U.K. (Tel: +44 171 253 4488; Fax: +44 171 430 0514; Email: aslib@aslib.demon.co.uk)

Machine Translation: Ten Years On
12-14 November 1994, Cranfield, U.K.

John Hutchins

In conjunction with Cranfield University, the Natural Language Translation Specialist Group of the British Computer Society organised an international conference to mark the tenth anniversary of their successful 1984 conference. The programme was designed to reflect the achievements of the past ten years and to help to stimulate further research and development in MT and MAT in the next ten years.

The opening speech by Yorick Wilks launched the conference with a typically stimulating overview of current trends in MT methodology, looking critically at the 'statistical turn', at the difficulties of international cooperation on resources for R&D, and at questions of evaluation. Three other speakers reviewed developments historically: John Hutchins surveyed the changes in MT methodology since 1984, with particular emphasis on the emergence of corpus-based approaches; Angeliki Petrits described the growth in use of the Systran MT system within the European Commission; and Peter Wheeler narrated his experiences with the Logos Corporation and as an independent MT post-editor and consultant.

A number of papers were concerned with fundamental theoretical issues of MT. Jörg Schütz (IAI, Saarbrücken) gave an account of his research on the integration of domain-specific terminology knowledge base and the powerful unification grammar formalism ALEP. Richard Morgan (Durham University) described the ambitious LOLITA project for building a large-scale independent knowledge-based NLP engine, and some small-scale demonstrations of its application to Italian-English translation. Galja Angelova (Bulgarian Academy of Sciences) reported on the joint development with Hamburg University of a knowledge-based aid (DB-MAT) for translating terminology from German into Bulgarian. Chadia Moghrabi (Université de Moncton) described her experiment on French and Arabic translation of cooking recipes to illustrate problems of text generation. Iris Höser and Barbara Rüdiger (Gesellschaft für Multilinguale Systeme mbH, Berlin) gave an instructive account of the difficulties of Russian-German translation based on experience with the VIRTEX system of the previous Deutsche Akademie der Wissenschaften and their subsequent research for the Sietec METAL system. Christian Boitet (Université de Grenoble) described his recent research on the 'dialogue-based' LIDIA system and suggested that interactively disambiguated ('self-explaining') documents would be less ambiguous source texts than those produced in controlled-language systems.

The problems of anaphora were the subject of three papers. Wilhelm Weisweber (Technische Universität Berlin) described the KIT-FAST project for German-English translation which involves the derivation of representations of text structure and text content, with particular emphasis on anaphoric relations. Ruslan Mitkov (IAI Saarbrücken) expressed the view that discourse relations had been too much neglected in MT research in the last ten years, and described his own impressive work on the identification of the antecedents of anaphors. It was a pity that Horacio Saggion and Ariadne Carvalho (UNICAMP, Brazil) were unable to be present to give their paper on anaphora resolution in a system for translating scientific abstracts from Portuguese.

Various aspects of the large Verbmobil project supported by the German Ministry of Research and Technology (BMFT) were the topics of papers by Susanne Heizmann, Bärbel Ripplinger and Scott McGlashan. The ultimate aim is a portable device for translating spoken German or Japanese into English in the context of business negotiations, and a number of research teams have been set up in German universities. Susanne Heizmann (Universität Hildesheim) described the research on modelling strategies used by human interpreters to achieve communicative goals in dialogue situations. Bärbel Ripplinger (IAI Saarbrücken) concentrated on the features required in conceptual representations for spoken language translation. Scott McGlashan (Univ. Saarbrücken) considered the semantic properties of a spoken dialogue system, drawing on experience with the SUNDIAL project for development in Verbmobil.

Although the conference papers concentrated primarily on experimental systems, there were also descriptions of working practical MT systems. Terence Lewis outlined the fundamentals of the sublanguage system developed in his company Hook and Hatton Ltd. for the translation of chemical engineering texts from Dutch into English. Subsequently, the system has been adapted to other restricted domains with impressive success. Svetlana Sokolova described and demonstrated the Russian-English STYLUS system marketed by her St.Petersburg company PROject MT Ltd., which also illustrated what can be achieved with relatively unsophisticated computational linguistic methods. In this respect, it was disappointing that another Russian, Mikhael Blekhman, was unable to travel to the UK in order to present his PARS system for Russian, Ukrainian and English MT. Finally, Aidan McGuire gave a demonstration of the EZ Japanese Writer for Windows, a tool enabling English speakers to create business correspondence in Japanese.

The user perspective was also not ignored. Ursula Bernhard (German National Research Centre for Computer Science) drew upon her eight years of experience as a user of Logos and Metal systems to account for the slow take up of MT by the translation community; and Veronica Lawson reported on the IAMT survey last year, which demonstrated the substantial increase in the use of MT systems in the past few years, particularly by multinational companies, but most notably by non-professional purchasers of the cheaper MT systems from Globalink and Microtac.

As this report illustrates, the Cranfield conference covered a wide range of MT topics; and it was the general wish of participants that another similar conference should be organised before the passing of another ten years. It is proposed to publish the full proceedings within the year.

Summer School on Computers and Translation
5-8 July, University of Vigo (Galicia, Spain)

Organised by Javier Gómez Guinovart (University of Vigo), the summer school was conducted in Spanish and Galician. On the first day, Victor Furundarena (AT&T Madrid)

described the MT activities of the AT&T corporation, and Santi Pons (IBM Barcelona) demonstrated the TranslationManager/2 workstation. On the second day, tutorials were given by José Fco. Quesada Moreno (Centro de Informática Científica de Andalucía, Sevilla) and Juan Albert Alonso (INCYTA Barcelona). The first tutorial ("La traducción automática: un reto informático") discussed the computational requirements and demands of MT, both in hardware and software; the second ("Qué es la traducción automática?") illustrated problems and limitations of MT systems from the perspective of the Siemens' Metal system. The third day included a talk by Joseba Abaitua (University of Deusto) on the present and future of MT, and a paper by Anxo M. Lorenzo Suárez and Javier Gómez Guinovart (University of Vigo) entitled "Terminoloxía informática", which described a glossary of computing terms in Galician, Spanish and English in the form of a hypertext Windows document. The final day of the summer school consisted of a presentation by Ana Aquilar-Amat Castillo (Univ. Barcelona) on the linguistic problems of MT lexicology and a discussion of technical translation of computer literature by Alfred Comin, editor of *Binary*, the Spanish computer magazine. Further information may be obtained from: Prof. Javier Gómez Guinovart, Universidade de Vigo, Facultade de Humanidades, Apartado 874, E-36200 Vigo, Galicia, Spain.

The LISA Forum
Mountain View California, November 7-8, 1994

[Edited extracts from a full report by *Michael Anobile* to be published in 'LISA Forum Newsletter'.]

Introduction: SeChang Oh from SunSoft Inc. welcomed the attendees; he noted that since 60% of SunSoft sales are outside the U.S., the company has many concerns related to I18n and L10n. At this point, Michael Anobile, Executive Director of LISA, talked about what he hoped would happen over the next two days, noting that there were two tracks and an emphasis on Asian issues.

Localizing for the Double-Byte Markets: Victor Tan (Director, Asian Localization, R.R. Donnelley Language Solutions). Victor spoke briefly about Beijing and the opportunities opening there and then segued into his presentation material. How does one go about localizing for the Asian market? First, Victor emphasized that one must understand the basic functionality of the application being localized and then also the cultural being localized for. In addition, one must understand the original coding and design of the application, etc.

Victor went on to talk about DBCS (double-byte character sets). He explained that because of the number of characters in the Japanese (7000 characters) and Chinese languages (perhaps 40,000 characters), that the scope provided by double-byte character sets is required. On the other hand, companies want to keep a single code base and not have separate code bases for different locales.

Localization Potential in the Pacific Markets: Siu-Ling-Koo (Managing Director, L&L Informatie). Siu-Ling-Ku noted that her country is a non-DBCS country. She asked: "Does it pay to localize? She answered that some issues that should be involved in this decision are

* Whether it is required by law
* What the possible revenues are (purchasing power)
* Difficulties involved
* Education of the population, etc.

She showed a graph showing the percentages of the world's population speaking certain languages; 20% of the world speaks the mainland version of Chinese. However, the possible revenue there is very small. In addition, education is comparatively low.

Siu-Ling said that somehow, one needs to look at language, purchasing power and educational levels, then factor in considerations like telephones per capita as another indication of technological sophistication, and then decide whether the expense might pay off.

Siu-Ling said that two versions of software are needed for the Malaysian and Indonesian markets because of the different derivations of the technical vocabularies: technical Malaysian brings in many English words while Indonesian brings in words from Dutch.

Translation Assistance Tools: Basic Requirements for English to Japanese Projects: Mineo Yamakawa. The first step must be string extraction and externalization; documentation includes software, online manuals, etc. Online formats for text include ASCII and Shift-JIS. Graphics with text include Windows metafiles, EPS, and other application specific formats. Document formats and styles are RTF and application-specific. The actual translation phases include literal translations with maximum information transfer, optimization of document readability, final reorganization of material in preparation for localization.

Dr. Yamakawa had a "wish list." Some of the list focused on kana-kanji conversions.. He noted that DOS/V and Windows J are stable products now... He is hoping that programs will be easily portable and that parts of programs can be combined. Graphics and text extraction should have features that are extractable (e.g., color from RTF files). What happens when the source text has been changed without letting everyone know? There should be some way of indicating to the localizer that this has happened without having to go back and do a check of every single bit of text.

Framework of a Standard Bidding Platform: Arthur Braunstein (AT&T Business Translations, Operations Director). Arthur Braunstein of AT&T to talk about work being done by LISA members on developing standardized RFQs. This issue was discussed at the last LISA Forum in Boston and it was agreed that a standardized format would be beneficial to both the seller and user of translations.

The Center for Software Development: Pamela Parrish (Executive Director, Software Development Center). The Center for Software Development is located in San Jose and is a collaborative effort of the City of San Jose, Novell and the Software Entrepreneurs Forum. Their mission is to grow software businesses by providing testing and porting labs, informational programs like educational series, and library services.

Japanization and On Screen Character Generation: CJK Font Support in PostScript: Ken Lunde (Adobe Systems, Inc.). CJK character sets include traditional and simplified Chinese, Japanese and Korean and Unicode/ISO 10646. Various encoding sets, used for information interchange, include ISO-2022, EUC and Shift-JIS for Japanese, and Big Five for simplified Chinese. Ken showed a matrix of code sets vs. character sets to show which code sets were rich enough to support which character sets.

Adobe offers five character sets under development. Adobe-Japan1-2 include 8719 CIDs; JIS X 0208-1990 supports NEC, Fujitsu, Apple and MS ext. And in addition, ISO 2022, EUC and Shift-JIS encoding support.

Panel Discussion: Terminology Ownership, Pricing and QA Responsibility. Moderator: Philippe de St-Maresville (Hewlett-Packard); Panelists: Maria Henriksen (Microsoft), Jose DeHoyos (Novell Corp.), Theodora Landgren (Bureau of Translation Services) Iko Knyphausen (Trados GmbH)

The session focused on Terminology ownership and Translation memory. The main issue was the right to reuse translation memories and terminology.

Major discussion items/viewpoints:

- - one cannot protect a single term unless trademarked

- - glossaries are compiled and can be copyright protected; compilations cannot be reused or resold

- - the work resulting from translation memories belong to the publisher (this position is currently under review by Microsoft legal department)

- - definition of translation memory suggests units of two or more sentences, uniquely translated and can be considered protectable, not single terms

- - what is ownership? copyright infringements, individual terms, product glossaries, translation memories

- - can vendors / publishers reuse translation memories from related texts? ... key issue - can a vendor who has invested in the tools and process, reuse the translation memories for other clients?

Interoperability Between Unicode and UTF on UNIX: Ik Kim (Senior Localization Engineer, MTS Software Localization, SunSoft).

1995 Language Engineering Reports: Rose Lockwood (Senior Consultant, OVUM Ltd., London). Rose said that the second LISA survey is about to be released. Because of the proliferation of data and the fact that language engineering has really come to accepted, OVUM has decided to do two reports next year. One is a straight marketing report and the other is a more technical evaluation of services and technologies available.

What is globalization and who uses the term? It seems to be used slightly differently in the U.S. and Europe. As for the need for globalization, some of the topics handled are a review of issues that drive the market for LE; globalization of products and services, etc. Rose noted that the current trend toward global markets, even aside from software, requires that people be able to accept the use of other languages and be aware of other cultural considerations.

Therefore, where are the applications for multilingual business information? What sectors does this include aside from software and IT? What about multilingual technical documentation and the fact that companies like Compressive have noted the trend and stuck a machine translation system (for French) on their system to facilitate global interchanges? ... One interesting note is that controlled English doesn't seem to be of much concern in this industry while it is important in others.

For systems, OVUM is looking at usability and ease of implementation; customizability (user-definable features); applicability to total translation process, language and domain coverage (language pairs, for example); language processing technology, including linguistic depth and translation memory; terminology acquisition (tools, single or inflected, etc.); dictionaries (mono/bi/multilingual, tagging information, etc.); output quality test and onsite testing (batch translation and translation memory tests); a small tool user survey.

Paving the Critical Path in Localization: A Manager's Guide: Pablo Ghenis (Localization and Engineering Services Manager, AT&T Business Translations). Pablo said his talk would focus on localizing legacy software rather than internationalizing from the inception of a software design... One of the problems involved is the diversity of skills required nowadays for a project that may even involve artistic talent (for voice-overs, graphics, etc.). Therefore, a manager cannot expect to own all the skills required to do a project: he must assume that he is dealing with competent professionals. Pablo recommended giving distributors specific review responsibilities such as glossary approval, and also to give

the distributors veto power at the same time. They should be getting early samples of translation for review and comment for fine-tuning before the entire manual set is completed and the distributor objects to the tone of the documentation.

UNICODE: The Business Case: Lloyd Honomichl (Engineering Manager, Novell)

Software Internationalization: Quality Assurance: Emmanuel Uren (Independent Consultant). Emmanuel said he wanted to talk about errors, tools to detect errors and error prevention techniques. He went through some definitions of internationalization and localization, quality and assurance.

Panel Discussion: Internationalization Issues Which Help Maximize Localization Efficiency and Management. Moderator: Shripad Patki, (SunSoft Kit and Technology Manager). Panelists: Ik Kim, Lloyd Honomichl, Ken Lunde, Pablo Ghenis, Pierre Cadieux. The panel dealt with the differences between I18n and L10n, how to handle different character representations, object handling, etc.

Ik said that an internationalized design can either be too generalized or that the interface is so specific that it will not work with some locales. Lloyd said he would rather think of markets rather than products. Pierre said that Windows 3 had been consolidated in locale blocks but typically Asian is seen as one case, European as another, and bidirectional as a third... Ik commented that English is no longer the language that is developed for first and that this makes it easier to handle simultaneous releases.... Lloyd said that localizing multimedia is a silly thing to do and shouldn't be done... He mentioned the localizing of encyclopedia for the Japanese market and noted that much of it would be irrelevant. He noted that content and context are closely related.

Operation Review: Gabor Ellman (Director of Product Translations, Oracle Europe). Gabor noted that 27 languages are supported by translations and are divided into Tier 1 and Tier 2 languages... In addition, many other languages are supported through character set support, etc. In addition, the applications are localized with regard to the expected sophistication of the typical end user of a particular product. Gabor addressed what he calls depth of localization, with GUIs; message and prompts; seed data (default data, setup values, etc.); and online help pages needing localization in addition to normal documentation. Documentation localization is not critical because it may or may not be used.

Gabor said that Oracle spends about US$15M on localization, apart from people who maintain run-time libraries, etc. This is mainly external vendor costs. The output is 20 applications localized into 20 languages. It includes 2000 screens, 2M source words of online help and 3.5M source words of documentation. For products other than applications: there are 0.2M source words into 25 languages with online help of 1.8 M words. Approximately 15M words are used, primarily for Asian languages. Gabor noted that paper media is being phased out; online documentation is being provided on CD-ROM...

DOCWARE Workgroup Computing for Localization: Rafik Belhadj-Kacem (Localization Manager, BULL-ILO). Rafik said that the goal of the workgroup, which is funded by the EC, is to reduce L10n costs and delays. He said that the technologies used by the group were not developed by the group but are those developed by others or by BULL. They are interested in client/server, archiving, automation, protected access, etc. In addition, controlled languages, MT and information retrieval and searches are being investigated. The main component of the project is the localization work station. it should be simple and easy to use and be secure. The other main component is multilingual electronic documentation systems, especially indexing. It is important to note that the indexer is human-assisted, not computer-assisted. The project size is small, 78 person-months, and should be finished by the end of 1995. Bull, CNRS, Linga, and The Open University are partners in this project.

MT at ANLP-94

At the *Fourth Conference on Applied Natural Language Processing* held at the University of Stuttgart from the 13th to the 15th October 1994, a large number of contributions related directly or indirectly to machine translation. Naoto Katoh and Teruaki Aizawa described the English-Japanese system under developed at NHK for the translation of economic news stories (ENTS) in a paper on "Machine translation of sentences with fixed expressions", a type of example-based approach which has shown considerable advantages over conventional rule-based methods. Ido Dagan and Ken Church contributed a paper on "Termight: identifying and translating technical terminology". *Termight* is a semi-automatic tool for professional translators to identify technical terms and their potential translations, which has been developed within the context of the statistical corpus-based research at AT&T Bell Laboratories. Akitoshi Okumura and Kazunori Muraki report on recent developments of the English-Japanese system at NEC in a paper entitled "Symmetric pattern matching analysis for English coordinate structures". Miguel Filgueiras (Universidade do Porto, Portugal) describes in his paper "A successful case of computer aided translation" a method for translating a mathematics textbook from Portuguese into English. Robert Frederking and Sergei Nirenburg ("Three heads are better than one") argue from experience of the Pangloss project that MT quality will improve if systems integrate a variety of methods instead of being limited to one single approach. Kozo Oi et al. describe some recent developments of the Transfer-Driven Machine Translation model for the ATR project in "Real-time spoken language translation using associative processors". A number of the poster sessions were also devoted to MT, e.g. another paper on the ATR project by Susumu Akamine et al., two papers on the NTT research by Satoru Ikehara et al. and by Kentaro Ogura et al., a paper on Toshiba MT research by Hideki Hirakawa et al., and a paper by Reinhard Schäler (University College Dublin) on MT evaluation. In addition, the ANLP-94 proceedings record a continuing growing interest in research on multilingual generation, with papers on the TECHDOC project (Ulm, Germany), the FoG project (Montreal, Canada), and the EXCLASS system at CoGenTex Inc. Other areas well represented are papers on text alignment, automatic tagging of large text corpora, lexical acquisition, and 'hybrid' symbolic-statistical methods in general. [Details of the published proceedings are given in 'Publications Received'.]

MT at other Recent Conferences

Research in NLP comparing and integrating rule-based and statistics-based methods has been a dominant theme in a number of conferences in the latter half of 1994. For this reason, the proceedings should be of interest to many researchers in MT, where 'hybrid' approaches have been the focus of intensive activity in recent years. An ACL workshop on 1 July at New Mexico State University was entitled "*The Balancing Act: combining symbolic and statistical approaches to language*". MT itself was the focus of two talks by Hiyan Alshawi ("Qualitative and quantitative designs for speech translation") and by Carolyn Penstein Rosé and Alex Waibel ("Recovering from parser failures: a hybrid statistical/symbolic approach") on the JANUS speech translation system.

From 14th to 16th September, the University of Manchester Institute of Science and Technology hosted an *International Conference on New Methods in Language Processing (NeMLaP)*, which was also devoted to statistical and symbolic approaches, with MT itself as the direct topic or major aspect of a number of papers: "Towards automatically aligning German compounds with English word groups in an example-based translation system" (Daniel Jones and Melina Alexa), "A natural language translation neural network" (Nenad

Koncar and Gregory Guthrie), "A full-text experiment in example-based machine translation" (Sergei Nirenburg, Stephen Beale, and Constantine Domashnev), "Self-organizing example-based machine translation: a prototype" (Patrick Juola), "The exploitation of parallel corpora in projects ET10/63 and CRATER" (A.M.McEnery et al.), "Direct parse tree translation in cooperation with the transfer method" (Yosihiro Matsuo et al.), "A paradigm for non-head-driven parsing: parametrized message-parsing" (Bonnie Dorr et al.), "Extracting semantic features for aspectual meanings from a syntactic representation using neural networks" (Gabriele Scheler), and "Some methods for the extraction of bilingual terminology" (Éric Gaussier and Jean-Marc Langé).

A workshop held by ISSCO in Geneva on 2-3 December was devoted to *Compound Nouns: Multilingual Aspects of Nominal Composition*. It was the third is a series, in which the first two had been held in Fontenay (1992) and Paris (1993). Given the focus of the workshop, most papers have some relevance to MT, and some considered translation problems directly, e.g.: "Relational adjectives in the translation from Germanic nominal compounds into Romance languages" (Marta Carulla), "Analysing and generating English compound structures for machine translation" (Chris Chambers), "English-Arabic machine translation of nominal compounds" (Zouhair Maalej), and "Traitement des unités lexicales complexes en traduction automatique" (Marie-Claude L'Homme).

Details of the published proceedings of these conferences and workshops will be found in the section 'Publications Received'.

ASSOCIATION NEWS

ASSOCIATION FOR MACHINE TRANSLATION IN THE AMERICAS

Letter from the President (Muriel Vasconcellos)

AMTA Comes of Age

Dear Colleagues,

With the success of its first conference, "Technology Partnerships for Crossing the Language Barrier" (AMTA-94, Columbia, Maryland, 5-8 October), AMTA not only joined the big leagues but also clearly validated its fundamental purposes. The turnout of 165 registrants from 13 countries surpassed expectations and reflected the wide range of members' interests and origins that is the hallmark of AMTA and its sister associations.

AMTA-94 may indeed have been the first exclusively MT conference to offer parallel sessions, and it was certainly the first non-Summit conference to offer such a rich variety of demonstrations, exhibits, technical presentations, tutorials, lively panels, and special interest group meetings. The extra effort entailed in implementing this format, which added a notable burden to the work of Program Co-chairs Eduard Hovy and Joseph Pentheroudakis, was well worth it. The conference was in fact able to attain its goal of offering "something for everyone" by providing timely and refreshing material of interest to researchers, developers, and actual and potential users with an impressive range of requirements – both those already with extensive MT experience and newcomers to the field.

The conference proceedings (edited by Joseph Pentheroudakis) were the final confirmation of AMTA's coming of age. This volume, containing over 30 refereed papers as well as summaries of the panels and demonstrations, is now selling briskly to people who could not attend. Also distributed at the conference was the report of the NSF-sponsored MT

Evaluation Workshop organized by AMTA in San Diego on 2-3 November 1992. These are our first two substantive publications, and we are very proud of them. (For information on ordering copies, see elswhere).

Not only was the conference a technical achievement and a financial success; it served to build a sense of community among MT-ites. It also gave visibility to MT as a technology: the conference received nationwide publicity in USA Today, which covered d'Armond Speers and his Klingon Language Analyzer.

The smooth and successful operation of AMTA-94 was due to the contributions of many. Program Co-chairs Eduard Hovy and Joseph Pentheroudakis worked unstintingly (and always cheerfully) for months before the event. The Program Committee, comprising a roster of 20 eminent MT specialists from around the world, were most helpful in preparing insightful reviews, offering suggestions to the authors, and working out differences of opinion among themselves. Extensive time and energy were contributed by Bill Fry, Exhibits Coordinator; Jane Zorrilla, AMTA-94 Registrar; and Debbie Becker, AMTA Focal Point, to ensure that everything ran smoothly.

One of the important developments at AMTA-94 was the formation of five special interest groups: MT Evaluation (Marjorie Leon and John White, Co-chairs), Interlingual MT (Eduard Hovy, Chair), Standards and Data Exchange (Alan Melby, Chair), MT on PCs (Chris Miller and Michael Tacelosky, Co-chairs), and The MT Lexicon (Bonnie Dorr and Joseph Pentheroudakis, Co-chairs). Each of the SIGs met and drew up an agenda. We expect them to organize workshops, prepare educational materials in their specialized areas, and develop initiatives that will further AMTA's purposes.

In the future we plan to hold the regional conference biennially in even years. The SIGs, in turn, will hold their events and do their more intensive work in odd years, and the results will feed into the next regional conference. That's the Master Plan, and it's now in place.

AMTA has also came of age in the sense that definitive Bylaws were ratified at the annual membership meeting held in conjunction with the conference. The first general election had been conducted by mail the month before, and the results were announced at the meeting (see the Secretary's report below). I would like to join the members at this time in recording our special thanks to the outgoing Treasurer, Roberta Merchant, who steered us through some very shallow waters when the Association first started and whose wise management now enables us to sail full speed ahead. Roberta remains on the Board and continues to be Treasurer of IAMT. I would also like to officially welcome our new Treasurer, Claudia Gdaniec, as well as Eduard Hovy, Vice President, and the other members of the newly elected Board: Bonnie Dorr, Mary Flanagan, Denis Gachot, Elliott Macklovitch, and Sergei Nirenburg. Scott Bennett remains as Secretary, and I will continue to serve as President for another two years.

Let me take this opportunity to wish all the members of AMTA and our sister associations a happy and prosperous 1995. It should be an exciting year for MT!

AMTA GENERAL MEMBERSHIP MEETING
October 7, 1994

Minutes of the previous meeting were approved as reported.

PRESIDENT'S REPORT: Muriel Vasconcellos reported the following points of interest: Membership is now 215 members. The conference was a big success, a big turning point for the AMTA. Attendance was 155 people; excellent set of papers. A new Yellow Book will be published

in early 1995. Members of the EAMT and AAMT will also be listed, if they desire. Debbie Becker is the Focal Point for the AMTA. She will handle association business as necessary.

BYLAWS: The revised bylaws were approved unanimously by the membership. Notable changes include: Membership categories are now: individual, corporate and institutional. Size of the Board has been reduced. Officers' terms are no longer staggered to allow a consistent Executive Committee and permit officers to run for other offices. Votes are now allowed by mail ballot.

TREASURER'S REPORT: Roberta Merchant reported the financial status of the AMTA with particular expenses and income from 1994. Of note were that the MTNI costs $1500-2000 per issue and that the conference made around $8000, which will be used as "seed money" for the next one. The AMTA should end the year with approximately $15000 for next year.

EDITOR'S REPORT: Joseph Pentheroudakis reported that three issues of the MTNI had been published in 1994, each averaged 48 pages. The EAMT issues are also published in the US, since the cost is considerably less here. Jane Zorrilla will help with layout and publishing matters. Back issues up to now are available from Joseph; 1994 and later issues will be available from Debbie Baker.

INFORMATION COMMITTEE REPORT: Jackie Murgida, Chair, reported some of the activities of the committee including: putting together "boilerplate" responses to frequently asked questions, doing bibliographies and referrals upon request and creation of a frequently asked question list. For the future they are considering a network service and putting a request for information in the MTNI.

ELECTION RESULTS: The following are the results of the election:

President: Muriel Vasconcellos
Vice President: Eduard Hovy
Secretary: Winfield Scott Bennett
Treasurer: Claudia Gdaniec
Directors: Bonnie Dorr, Mary Flanagan, Denis Gachot, Elliott Macklovitch, Sergei Nirenburg, and Roberta Merchant.

SPECIAL INTEREST GROUPS: 5 special interest groups (SIGs) were formally established at the conference:

Standards: Alan Melby, Chair.

Interlingual MT: Ed Hovy, Chair. 19 people at first meeting. Plan is to collect information and have short meetings in conjunction with conferences.

Lexicon: Joseph Pentheroudakis, Chair.

Evaluation: John White, Chair. 20 people at first meeting. Plan to have a quasi-monitored list of issues via eric@cmt by November 18.

PC translation: Chris Miller and Mike Tacelosky, Chairs. 6 people at first meeting. Plan is to produce an information guide, an evaluation guide and a guide for buying PC-based systems and to have an AMTA forum on ComupServe or such a network.

Respectfully Submitted, Winfield Scott Bennett (Secretary)

Brief News from AMTA Board Meeting

The Board of the AMTA met on October 8, 1994. During that meeting, many of the preexisting committees, such as the Fund Raising Committee, were eliminated (e.g., the Fund Raising Committee) or folded into another committee or one of the Special Interest Groups (e.g., the Evaluation Committee has become the Evaluation SIG and the Public Relations Committee has been subsumed under the Information Committee). The Membership Committee remains as it was.

The Board would like to let AMTA members know that the Nominations Committee needs more people in order to represent all sectors of our organization in fulfilling its mandate.

Finally, the Board approved the following annual dues structure:

Membership dues:

Individual:	$60
Institutional:	$200
Corporate:	$400

Reinstatement Fee:	$25
Newsletter Subscriptions:	$75

Advanced Translation Studies Center (ATSC), Saarbrücken

[From: LINGUIST List, 5-1421.]

The Department for 'Applied Linguistics, Translating and Interpreting'(Fachrichtung 8.6 'Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen') at the University of Saarbrücken belongs to the established university institutes in the Federal Republic of Germany offering degree programmes for future translators and interpreters. The Department has now founded the Advanced Translation Studies Center.

The Center offers one-week courses on the state of the art in translation/interpretation as it is taught in the degree courses for translators and interpreters. The courses on the translation of LSP texts, culture specifics in texts, multilingual terminology management, machine translation and interpretation reflect the School's areas of specialization. International guest speakers will provide a broader international context. Target group: (prospective) students and teachers in translation/interpretation, practitioners in the field and all those who are interested in obtaining a quick overview of the academic discipline. In spring 1995, the following courses will be held:

I. Languages for special purposes (LSP) and translation.
II. Culture and translation
III. Multilingual Terminology Management
IV. Machine Translation
V. Interpretation

Information on all courses

Language: English (with reference to German)

Place: University of Saarbrücken, Fachrichtung 8.6, Postfach 15 11 50, D-66041 Saarbrücken, Im Stadtwald, Bau 4

Length: 1 week, 9.00 - 13.00 h, July/August and February/March

Academic Responsibility: Prof. Dr. Heidrun Gerzymisch-Arbogast (Tel: +49 681 302-4246; Fax: +49 681 302-4440)

Costs: DM 1500 per block of instruction

For further information please contact: Dipl.-Dolm. Jessica Pfeil, University of Saarbruecken, Fachrichtung 8.6, Postfach 15 11 50 (Tel: +49 681 302-4248; Fax: +49 681 302-4440; Email: Jessica@dude.uni-sb.de)

Reseau Informatique et Linguistique (RIL), Belgium

[From ELSNET list]

The Minister of the Walloon Region of Belgium in charge of Research and Technological Development, Mr Albert Lienard, who always expressed his interest and his support for the theme of Linguistics and Language Engineering, wishes that the Research and Development centres of the Region continue to actively participate to the efforts in that domain.

As a consequence, the Ministry called for a consulting mission consisting in the creation and animation of a network of competence centres in Computer Sciences & Engineering and Linguistics, on behalf of the Region.

The aim is primarily to foster collaboration in the field of Language Industries and Computational Linguistics, in view to build projects, to seek for funding, and to exchange

information between universities and other higher education institutions, industry and users, both within the Region and in relation with other regions and countries.

The network 'Reseau Informatique et Linguistique', or RIL for short, will have a coordination cell as a focal node, the present address of which is in Louvain-la-Neuve.

The RIL network is not an additional structure which would somehow compete with existing structures such as the international research networks, or other bodies with a larger scope, such as the ACCT and its network of Language Industries Observatories (among which is included the OWIL, Observatoire Wallon des Industries de la Langue). It should rather be viewed as an action to facilitate the information flow between the actors of the domain, both within the Region and with other regions and countries, in Europe and elsewhere.

The interfacing function offered by the RIL network should allow to more easily and efficiently get in contact with centres wishing to establish partnerships to prepare and propose collaborative R&D projects.

The envisioned activities cover fundamental research as well as technological development, but also care for the expression of needs, in terms of products or services, the encouragement to the creation of new companies, the commercialisation, and the utilisation of research results.

Themes in computational linguistics include:

- fundamental research in computational linguistics, and
- applied research for the creation of resources, methods, tools and systems:
- resources: corpora, lexicons, grammars;
- speech recognition and synthesis;
- written text analysis, understanding, generation, synthesis;
- multilinguality: translation, multilingual texts;

Themes in language technology and engineering include:

- speech: vocal command, dictation, rereading, vocal annotation;
- written text: writing aids, verification, correction, version comparison, updating, annotation, hypertext links;
- translation and multilinguality: multilingual writing, MT or MAT, translation aid tools;
- document: creation, distribution, archiving, indexing, retrieval, annotation, integration of text, graphics, and sound, navigation systems;
- telecommunications: document exchange, teleworking for groups and individuals;
- person-machine interfaces: integration of technologies for the person-machine communication, data bases interrogation in natural languages;

as well as specific applications, such as:

- computer assisted learning and teleteaching;
- applications for developing countries;
- applications for the disabled and elderly;
- and so on.

Any interested person or organisation is very welcome to contact: Thierry J. van Steenberghe, RIL, UCL=University of Louvain, Department of Computer Science, Place Ste Barbe 2, B-1348 Louvain-la-Neuve, Belgium. (Tel: +32 10 47 3150 [or 2081]; Fax: +32 10 45 0345; E-mail: tvs@info.ucl.ac.be)

PEOPLE ON THE MOVE...

Muriel Vasconcellos, President of AMTA, has moved to:
2060 Reed Ave., San Diego, CA 92109.
Her email (CompuServe) address remains:
71024.123@compuserve.com

Colin Brace, editor of Language Industry Monitor, has a new email address: colinb@ibm.net

The new EAMT regional editor for MT News International is:
Dr. *Jörg Schütz*,
Institute for Applied Information Sciences (IAI),
Universität des Saarlandes,
Martin-Luther-Str.14,
D-66111 Saarbrücken,
Germany.
His fax number is: +49 681 397482; and his email address: joerg@iai.uni-sb.de

SYSTEMS and PRODUCTS

Latest version of ASTRANSAC from Toshiba Corporation

[Extracts from AAMT Journal no.7, June 1994]

ASTRANSAC was put on the market for AS-series (Toshiba's EWS) in 1986; ASTRANSAC for SunWS (individually purchasable software) is available since 1993. New version of ASTRANSAC has just been released.

Features of ASTRANSAC

* Translation speed. The number of translation words per hour is about 60,000 words (when using AS4080/SPARCstation 10).

* 100,000 grammar rules and the semantic transfer method.

* Technical term dictionaries and user-defined dictionaries to cover wide range of fields, such as information processing, machine, politics and economics,and chemistry. A maximum of 200,000 words can be registered into a user-defined dictionary. It is also possible to create more than one user-defined dictionary.

* User customization function. Setting up parameters (such as dictionaries and expressions to use) to understand the original text written in users' style, and to translate it into desired style with precision and homogeneity is simple in ASTRANSAC.

* OCR function to directly read manuscript. High recognition rates of 99.7% for English and 99.5% for Japanese achieved. The automatic layout analysis understands complicated layouts and vertical writing.

* Excellent man-machine interface. ASTRANSAC's interface is based on OPEN LOOK, the standard GUI of the industry. The direction of translation can be switched between Japanese to English/English to Japanese, with one button.

Features of latest version:

* Preediting support function for Japanese to English translation. The preediting, which is indispensable to japanese to English machine translation is now half-automated. Items to be focused on during preediting, which were only explained in technical manuals for users to find manually in previous version, are detected automatically, and

candidates for re-writing the detected expressions, together with guidance messages are displayed to the user. The items detected now are: long sentences, compound nouns, long hiragana words, un-registered words, redundant expressions, connectives, negative expressions, passive voice, and expressions registered in the "unsuitable expression dictionary". This preediting support function shortens time for preediting and improves translation quality.

*Dictionary editing function for English-Japanese translation. Setup of attributes or priorities to the content of dictionaries or new entries to dictionaries can be performed easily in a window. This function allows users to register complicated attributes, hence the quality of translation will be bettered.

*Translation for DOCMaker3 [a Toshiba product]. MIF document, the data for a highly sophisticated DTP software, DocMaker3, can be translated by ASTRANSAC. The layout of the text is reserved when the document created by DocMaker3 is translated, the result is saved in MIF format, and the MIF data is opened by DocMaker3. Of course, the same is true for documents of DocMaker. The time needed for postediting is shortened by this function.

* Interface with File Manager. Document can be registered by Drag & Drop of the File Manager (DeskSet of OpenWindows). The document registration became simpler.

For further information: Computer Division, Toshiba Corporation. Tel: +81 3 3457 2725

Bravice J/E ver. 5.1 by MT Laboratories Co., Ltd.

[Unedited extracts from AAMT Journal no.7, June 1994]

Bravice J/E ver.5.1 is the latest version of the world first Japanese into English translation software which was originally launched in 1984. The software has two ways of translation, bulk and spot. In the case of bulk translation, it translates a Japanese text prepared by using the built-in word processor. The user can select the whole-text translation, the blocked-text translation or the one-sentence translation. The size of the input file for translation is not limited and up to 16 files can be translated in sequence. Also a standard text file is available as an input even if it was made using another software. Spot translation is convenient when the user wants to translate few sentences quickly. The user doesn't need to make sentences in advance with a word processor, because English is output immediately by typing Japanese (maximum 400 characters at a time.)

The basic dictionary contains 60,000 words, and there are optional dictionaries which contain technical terms for 31 different fields. Moreover, up to 40,000 additional words can be easily registered. Even when it comes to selecting word attributes, which is very important in machine translation, the user simply chooses from a menu. An attribute is such information of words as "noun that means a person's name" or "uncountable" and so on. It enables our system to judge correct usage of those functional words such as relative pronouns or prepositions, by registering them precisely. Therefore, we can say that dictionary registration is indispensable to managing translation software. Also it is possible to register expressions in addition to words.

We need to input a concise and clear sentence in order to get a good English sentence. This is called "pre-editing". The checker function does this pre-editing automatically to a certain extent. It can check unregistered words in advance because it detects them at the same time. An easy-to-use word processor is built-in. The user can install Front End Processor they want. The user can operate translation and dictionary registration, dictionary searching in this

word processor. The user can operate these easily without being conscious of the translation software.

Features:

Software: Japanese DOS (or DOS/V)
Hardware: 20MB HDD is necessary
Price: Bravice J/E ver 5.1 = ¥190,000
Option dictionaries (from 31 fields) = ¥12,800-¥45,000
User support: Exclusive telephone/FAX
For inquiry, please contact:

PACIFIC EYE Inc.
Shin Osaka GH Bldg. 7F
6-9-20 Nishi Nakajima
Yodogawa-ku, Osaka 532 Japan
Tel: +81 6 885 1871
Fax: +81 6 885 0720

PC-Transer/ej for Windows: Nova's new English-Japanese system

[Unedited extract from AAMT Journal no.8, September 1994]

Our company had offered a real machine translation system for personal computers. The following group of products are quite supported by a lot of users.

* PC-Transer/ej: An English-Japanese machine translation system for: Windows, Macintosh, and MS-DOS machines including NEC PC98 series and DOS/V machines.
* PC-Transer/je: A Japanese-English machine translation system for: Windows, Macintosh, and MS-DOS machines including NEC PC98 series and DOS/V machines.
* SuperTranser/ej: An English-Japanese machine translation system which corresponds to patent documents for: MS-DOS machines including NEC PC98 series and DOS/V machines.
* Technical word dictionaries for the above systems as:
  E-J dictionary = 18 technical fields
  J-E dictionary = 15 technical fields.

Among the above, Windows version are new products which have joined the market in June 1994, with new translation support functions added, compared to MS-DOS versions and Macintosh version. I will introduce mainly the new functions of Windows versions here.

1. Adoption of step translation method. Our company developed "the step translation method." This method defines phrases in a sentence with complicated structure and complex modification. A more accurate syntactic analysis is achieved by an easy operation. Originally Nova' Japanese-English MT systems have been supported with this function, and its English-Japanese MT systems have its function as well now.

Example: (Original) *It outlines the minimum hardware and software requirements necessary to install and use Visual C++.*

The step translation leads a correct translation result.

1) Translate the part "the minimum hardware and software requirements" specifying as the noun phrase.

2) The part "the minimum hardware and software requirements necessary to install and use Visual C++" as a noun phrase.

3) Execute the whole translation, using the result of 1) and 2).

(Step translation result) *Sore ha, Visual C++ wo settchisite, sosite siyou suru tame ni hitoyouna saisyougen no hardware to software hiuyou jouken wo gaisetu suru.*

2. Equipped with thesaurus dictionary as standard. Info Soft Co.,Ltd. (150,000 words) is installed in a Japanese-English system as standard. In addition to selecting an English equivalent of the translation result with the basic function, the system makes the user select English synonyms from the thesaurus dictionary.

For instance, the first English equivalent for *honyaku suru* is *translate*. When referring to the meaning of "translate" in the thesaurus dictionary, it is given "to express in another language while retaining the original sense." A lot of synonyms are then displayed on the screen: "convert, decipher, decode, interpret, transcribe, transliterate." And the user can adopt the exact synonym for the translation.

3. Function to reflect former translation results. This function is supported to both English-Japanese translation and Japanese-English translation. For instance, this is useful for the translation of the revised version of a manual. Referencing Transer's output file, a user can easily know which parts of the new manual are changed from the original and should be translated. The user doesn't have to translate the same sentences repeatedly, even if they are edited. The user can make files just for reference as well. Therefore, it is able to use often used expression as the translation sentence.

4. Translation and editing of plural files. The plural files can be opened at the same time, and each file can be translated and edited. (Four files the maximum.)

5. Translates by the background processing. Two or more files are background processed, while not displayed on the screen. Naturally, a more high-speed translation can be done.

6. Server client method is adopted. This Windows version changed the system composition to the server client method. The "translation server" as the translation execution port, and the "client" as the user interface are separated. Thus, the operation similar to that of the server client on the network is achieved with one personal computer.

Besides allotting the function of the development side, the server client method will have the translation function be easily applied while using other application software in the future. For instance, use from the window applying word processor software, table calculation software and so on are assumed.

The new functions of the PC-Transer/ej for Windows version were introduced truly briefly above. For details, please inquire at: System Marketing Division, Nova Co.,Ltd., Rm 301, Suzusho Bldg., 23 Araki-cho, Shinjuku-ku, Tokyo 160 Japan. Tel: +81 3 3351 3356; Fax: +81 3 3351 5766.

TransLand Mac edition
Japanese-English translation software
by Brother Industries Ltd.

[Extract from AAMT Journal no.8, September 1994]

We have developed a Japanese-English translation software for in-company use. as a result, "TransLand PC98(MS-DOS) edition" (Japanese-English translation software), which is intended for personal use, was on the market last June (1993). Its translation quality and low price have been highly evaluated among the users.

Recently, the user interface of personal computers is changing to GUI, such as Mac OS and Windows. In this March [1994], we have released TransLand Mac edition so that more user can enjoy the easier use of our translation software.

Features

* Improved Translation Quality. TransLand PC98(MS-DOS) edition was improved as follows for Mac edition:

Better translation when Japanese with auxiliary, and Japanese particle *wa*.

Better translation of imperative sentences

More precise recognition of noun phrase

* Default Setting. The following default setting can be made so that the user can obtain translation closer to what they want:

Default setting of subject when there is no subject in the Japanese original

Default setting of article

Imperative or Passive can be set when there is no subject

Use of user dictionary can be chosen

If a word is taken from the user dictionary, the word can be marked in the translated sentence

Translation time can be set up to 120 seconds

* Improved User Interface. As Japanese original and translated English are shown side by side on the screen, correcting and re-translating the Japanese, and editing the English are done easily. Up to 9,999 sentences can be translated at a time. And, selected sentences in a document can be translated.

* User Dictionary Registration. Registration or modification to the user dictionary can be done on an independent window. On-line help is shown by just moving the mouse cursor to an input item.

* Price: ¥69,800

For information: Takamine, Business Machines Division, Brother Sales Ltd. Tel: +81 52 263 5855

PENSEE for Windows

[Extracts from AAMT Journal no.7, June 1994]

PENSEE for Windows is an easy to use JE/EJ machine translation software product designed specifically to run on a personal computer with word processor.

Features:

* As easy to use as any familiar Windows applications

* High quality translation. "Deep case grammar" analysis, a refinement of AI "case grammar", resolves "deeply embedded" grammar in long and complex sentences.

* Extended system dictionary for improved output. PENSEE for Windows is enhanced with business and economics dictionaries, for a total of 150,000 words for J-to-E, and 130,000 words for E-to-J. Optional dictionaries area available for 7 more fields; each field adds 100,000 words.

* Fast execution. PENSEE for Windows translates up to 15,000 words per hour.

* Flexible selection of user dictionaries. Up to 16 user dictionaries may be prioritized for automatic access.

* Easy manipulation of unknown words files. Unknown words can be culled from the document and filed in CSV format for editing. Registration to user dictionary is then done with the dictionary management function.

* Executes from MS-Word. Use MS-Word built-in Macro function while in MS-Word to execute.

* Enable user dictionary management with MS-Excel. Make, modify and manage user dictionaries with MS-Excel built-in Macros.

* Interfaces with other Windows applications (OCR, DTP, etc.); transfers texts via Clipboard.

Specifications:

Translation languages: Japanese to English, English to Japanese
Translation engine: PENSEE
Hardware: machine executable for MS-Windows 3.1J (Japanese)
Memory required: 8MB or more
Hard disk required: 40MB or more
OS etc.: Windows 3.1J
System dictionary: JE approx. 150,000 words; EJ approx. 130,000 words
Special dictionaries (optional):
JE approx. 70,000 words; EJ approx. 70,000 words
7 fields: electric, information processing, medical, machinery, chemical, etc.
User dictionary: unlimited
Translation speed: approx. 15,000 words/hour
Translation option: JE standard/MEIREI; EJ standard/KEITAI
Executable software:
MS-Word 5.0 (executing translation)
MS-Excel 4.0 (managing user dictionaries)
Prices
PENSEE for Windows: Y 198,000
Special dictionaries (7 fields packed): Y48,000

For more information: Oki Software Co.Ltd. Tel: +81 3 3454 7831

ATLAS for Windows

[Extract from AAMT Journal no.8, September 1994]

Fujitsu released its English-Japanese and Japanese-English machine translation system "ATLAS/WinV1.0", which works on the PC Windows environment, in May 1994.
This system has the following features.

1. High translation quality.
2. Twenty-four technical dictionaries with 1.2 million entries.
3. Variable translation styles.
4. OASYS/Win compatibility.
5. A tuning mechanism which allows the user to adjust the translation process.
6. Compatibility with any Windows 3.1 application.

Product system.

ATLAS/Win consists of the following four products:

1. "ATLAS/Win EJ V1.0" English-Japanese machine translation system.
2. "ATLAS/Win JE V1.0" Japanese-English machine translation system.
3. "ATLAS/Win Bidirectional Translation Option V1.0". This option enables ATLAS/Win EJ or ATLAS/Win JE to produce both E-J and J-E translations.
4. "ATLAS/Win Special Term Dictionary V1.0". This product contains 1.2 million entries covering 24 different technical fields which can be specified by the user in accordance with translation requirements.

The ATLAS/Win system.

Since launching the world's first commercially available machine translation system in 1984, Fujitsu has striven to improve translation quality while down-sizing and adapting to the open system environment. ATLAS/Win is the product of this process and incorporates the following features to produce high quality translations.

1. Semantic processing method. There are several different machine translation methods currently in use. ATLAS uses the Semantic Processing Method, which is considered one of the most advanced. It is superior to the Syntactic Transfer Method which replaces the structural pattern of the sentence. ATLAS/Win analyses the concepts of the input sentence using dictionary and semantic rules. From the results of the analysis a second set of dictionary and semantic rules are used to construct the translation. This method produces a translation with a high degree of sophistication and accuracy. For example:

A. Original: *Nihongo wo eigo ni keisanki de honyaku suru*.

Translation: *The machine translates Japanese into English*.

From the results of its analysis ATLAS/Win selects *keisanki* (computer) as the subject even though the original sentence does not have a subject.

B. Original: *Nihongo wo eigo ni te de honyaku suru*.

Translation: (*S) *translates Japanese into English by hand*.

Even though this sentence has the same structure as example A, ATLAS/Win recognises the words *te de* as indicating the method and not the subject and gives the adverb phrase *by hand* in the translation.

2. Basic Dictionary. ATLAS/Win has a basic dictionary with 130,000 entries. Information necessary for translation, such as spellings, inflections, parts of speech, semantic attributes, and English or Japanese equivalents, is stored in this dictionary.

3. Technical Term Dictionary. The Technical Term Dictionary actually contains 24 different technical dictionaries covering fields as diverse as information processing, medicine, finance, engineering, physics and atomic energy,and law. Users can choose and change dictionaries according to the nature of the translation being undertaken at the time. Compare the following translations.

Original: Open windows

Translation <Basic Dictionary only> *mado wo akete kudasai*

Translation <Using the Information Processing Dictionary> *uindou wo oopun shite kudasai*.

These special term dictionaries can be used for both English-Japanese and Japanese-English translation.

4. Tuning Mechanism. ATLAS/Win allows the user to 'tune' the translation process to meet the specific needs of the translation project at hand by using the following.

i. User Dictionary. New words or additional translations for current words can be registered in the User Dictionary. These words can then be used by ATLAS/Win for both English-Japanese and Japanese-English translations.

ii. Variable Translation Style. The user can select the most appropriate sentence style, casual or polite, active or passive, and imperative or non-imperative. For example:

A. Original: *I speak English*

Translation <casual style> *watashi wa eigo wo hanasu*.

<polite style> *watashi was eigo wo hanashimasu*.

B. Original: *buhin wo torihazusu*.

Translation <active voice> (*S) *detaches parts*

<passive voice> *Parts are detached*

<imperative> *Detach parts*

Translation format.

ATLAS/Win has three translation modes which make it possible to translate material from a range of sources, in small or large quantities.

1. Dialogical Translation. In this mode two windows, one for the original text and the other for the translated text, are displayed next to each other, either vertically or horizontally. The user types the original material onto one window and ATLAS/Win will generate the translation, line by line, on the other. This format makes comparison and cross-checking easy and the translation can be post-edited if necessary. This mode is especially convenient when writing letters or reports and it is possible to continue inputting sentences while the translation process takes place.

2. Batch Translation. This method translates MS-DOS and OASYS/Win text files directly without the need for conversion of file format. If large amounts of printed material are to be translated, OCR software can be used to save time.

3. Clipboard Translation. This mode allows the user to cut and paste text from any windows onto a special clipboard where it can be translated. The resulting translation can then be cut and pasted back onto the original window or onto any other window. This makes it possible to translate any text on an application compatible with Windows 3.1.

For further information: System Sales Promotion Section, Fujitsu Ltd., Marunouchi Center Building, Marunouchi 1-6-1, Chiyodaku, Tokyo, Japan. Tel: +81 3 3216 3211.

News from Logos

[Press releases]

Eurolang Optimizer for Logos

EUROLANG and LOGOS Corporation are completing the development of the first translation product that combines Translation Memory techniques with Machine Translation technologies. This new client/server based solution will combine EUROLANG OPTIMIZER with the LOGOS Machine Translation engine. Word processors and desktop publishing systems will be fully integrated with EUROLANG OPTIMIZER's and LOGOS' product line.

Easy to Use
Easy to Understand
Compelling Economics
State of the Art Machine Translation

EUROLANG OPTIMIZER for LOGOS is designed to give users maximum control over the translation process. users will see a window with clear and easy-to-read translations proposed by the LOGOS translation engine, sentence-by-sentence. They can quickly decide to accept the proposed translation, if necessary, post-edit it or translate manually.

EUROLANG OPTIMIZER for LOGOS offers ease-of-use assistance providing:

Perfect Matches. Sentences that have already been translated and which are found in the Sentence Dictionaries

Fuzzy Matches. Sentences that correspond approximately and which are found in the Sentence Dictionaries

Known Technical Terms. Technical terms found in the Term Database

Logos Translations. Sentences proposed automatically by the LOGOS machine translation engine

EUROLANG OPTIMIZER for LOGOS is available in multi-user configuration starting with 5 users at a cost of US$ 6,999/user.

Addresses: EUROLANG, 2 rue Louis Pergaud, F-94700 Maisons Alfort, France (Tel: +33-1-45 13 05 00; Fax: +33-1-45 13 05 59)

LOGOS Corporation, 200 Valley Rd, Suite 400, Mt Arlington, NJ 07856, USA (Tel: +1-201-398 8710; Fax: +1-201-398 6102)

LOGOS GmbH, Mergenthalerallee 79-81, D-65760 Eschborn/Ts, Deutschland (Tel: +49-6196-59 03 0; Fax: +49-6196-59 03 15

LogosClient for Windows Version 1.0

The LogosClient product is designed to support access to the Logos Translation Server software in a corporate local area network. This access can take place from the menus of the most popular word processing software packages in Microsoft Windows without further Logos application specific knowledge.

LogosClient is a Microsoft Windows based application program written in Microsoft Visual Basic for a standard IBM compatible PC.

LogosClient uses a mailing environment, connected via a LAN to the UNIX system, which runs the Logos server-based translation software. Specific macro programs perform the integration into the leading word processing software packages.

After LogosClient has been installed on your system, all its specific functions and network connections are activated automatically without user intervention.

Language-specific versions for an English and/or German working environment are available in the current version of LogosClient.

Logos functions accessible from LogosClient:

- Translation
- New word search
- Noun phrase search

Logos translation parameters defined at LogosClient:

- Source language selection
- Target language selection (single and multiple)
- Logos dictionaries by subject matter codes (up to 5)
- Customer-specific dictionaries (up to 5)
- Pre-editing environment

The following word processing software packages in their English and German versions under Microsoft Windows are supported by LogosClient:

Lotus Development Corp.	AmiPro version 3.0 and 3.01
Microsoft Corp.	Word version 2.0 and 6.0
WordPerfect Corp.	WordPerfect 5.2 and 6.0

The following software is required in order to run LogosClient:

Microsoft Corp.	Windows version 3.1
Sun Microsystem Corp.	PC/NFS version 5.0

The following minimum hardware is required to run LogosClient:

4 MB RAM minimum
2 MB additional free disk space
3.5" (1.44 MB) local floppy drive
Mouse
PC connected via LAN to the Logos Translation Server.

IBM launches LMT on the market

[Press release]

Automatic Translation with *Logic-Programming Based Machine Translation*

- high quality translation of technical texts
- language pairs: German-English, English-German
- integration with comfortable translation tools: IBM TranslationManager/2 and Keck MemCat
- state-of-the-art technology: knowledge-based architecture, logic programming, language-independent shell, powerful unification grammar
- uniform administration of all lexical and terminological storage via TransLexis
- user-friendly lexicon interface.

Technical data:

The machine translation system *Logic-Programming Based Machine Translation* runs on a single workstation as well as in a client-server configuration. The system can be used in a stand-alone environment as well as integrated with the translation tools IBM Translation Manager/2 and Keck Software's MemCat. The machine translation system *Logic-Programming Based Machine Translation* can also be run under IBM VM/ESA and IBM AIX.

Necessary hardware/software:

Single work station

Software:	IBM OS/2 2.1, IBM DB2/2, IBM Prolog/2 Runtime Facility
Hardware:	16 MB main memory, 20 MB spare hard disc space plus the necessary hard disc space for the terminological database, 386 processor or higher

Client/server version

Software:	IBM OS/2 2.1, IBM DB2/2 for client/server operation, IBM Prolog/2 Runtime Facility, IBM Communications Manager/2 or IBM LAN server
Server hardware:	16 MB main memory 20 MB spare hard disc space plus the necessary hard disc space for the terminological database, 386 processor or higher
Client hardware:	16 MB main memory 1 MB spare hard disc space, 386 processor or higher

Our service offer encompasses a broad spectrum of different services including:

- conversion of existing lexical and terminological storage
- development of new lexical data
- system installation and adaptation
- implementation of additional customer requirements
- training.

Contact: Ernst Nagel, IBM Deutschland Informationssysteme GmbH, WT Marketing, P.O.Box 10 30 68, D-69020 Heidelberg, Germany. (Tel: +49-6221-59-4281; Fax: +49-6221-59-3500.)

Price list

The machine translation system *Logic-Programming Based Machine Translation* is available as a single workstation under IBM OS/2 and in a client/server configuration as well as a mainframe system under IBM VM/ESA. The terminology management system TransLexis is included as a component.

IBM OS/2 single workstation	DM 3,500
IBM OS/2 client server version	DM 17,500
IBM VM/ESA mainframe version	DM 70,000

The prices shown here are for the translation system Logic-Programming Machine Translation with one source and target language: either German-English or English-German. The price for the second translation direction amounts to:

IBM OS/2 single workstation	DM 500
IBM OS/2 client server version	DM 1,000
IBM VM/ESA mainframe version	DM 1,000

The creation of the terminological data for the translation system is offered for the following price:

3,000 words	DM 10,000

The prices for training amount to:

At IBM Heidelberg per participant per day	DM 660
On-site training per day	DM 3,450

LMT for personal computers to come in 1995

It was also announced at a press conference on 6th October 1994 during the Frankfurt Book Fair that in cooperation with the electronic publishers von Rheinbaben & Busch Verlag (Munich, Germany) a version of LMT will be launched during 1995 to run on personal computers. The product - to be called *Personal Translator PT* - will require a 386 processor or higher (486 is recommended), running under Windows 3.1, with 8 MB RAM and 30 MB free hard disc space. It will translate in both directions (German-English and English-German) and include a dictionary of ca. 80,000 words. Facilities for an additional dictionary for nouns and personal names will be provided. The price is expected to be about DM 198.

At the same time, a more advanced version *Personal Translator PT plus* will be marketed, with a larger dictionary and facilities for several user dictionaries for different subject domains, which will allow adjectives and verbs to be entered as well as nouns and personal names. In addition, the advanced version will support texts in Word for Windows (2.0 and 6.0), it will provide a translation memory, and it will be capable of identifying the antecedents of pronouns. The expected price for Personal Translator PT plus is DM 498. For further information contact: v.Rheinbaben & Busch Verlag, Electronic Publishing, Georg-Hallmaier-Straße 2-4, D-81369 München, Germany (Tel: +49 89 7237777; Fax: +49 89 7238758)

News from Sietec

Colin Brace

[Extracts from *Language Industry Monitor* no.22, July-August 1994]

For many years the linguistics team at Siemens toiled away in comparative quiet, turning Metal from the promising German-English university prototype Siemens acquired from the

University of Texas in 1980 to one of the world's best MT systems, one that enjoys a small but growing contingent of faithful users.

These days, however, things are perceptively changing. The group, now known as Linguistic Systems, has been moved from Siemens to Sietec Systemtechnik, the long-time Austin development site has been closed, and the development team has completed the arduous task of porting the Metal system from the now obsolete Symbolics LISP machine to the Unix environment, making it easier to develop new languages pairs. And Metal is now being more aggressively marketed, both in Europe and in North America. At Sietec Systemtechnik, a systems integration and service group within Siemens-Nixdorf (SNI), Metal and its progenitors are closer to prospective customers and further away from R&D environs, where budgets are exceedingly tight these days. Like many of Europe's venerable IT companies, hard times have fallen on SNI; it racked up losses of DM600 million last year; however the Sietec Systemtechnik wing is profitable.

With no immediate prospect of directly recuperating the estimated DM50 million invested in Metal through sales of the system, Sietec is now actively looking for other ways to exploit the Metal technology. Says Thomas Schneider, director of the Linguistic Systems group, "we're now setting our sights beyond the small and very difficult translation market." Sietec would like to see the Metal engine deployed for a variety of language processing applications, from high end full-text information retrieval, message routing, and fact extraction systems, to end-user tools, such as grammar and style checkers, extended spelling checkers, and translation memories. With the substantial experience of SNI in large systems development behind it, the linguistics group is in an excellent position to develop custom-built NLP systems for large customers, and Schneider hints that negotiations are well under way with one very large prospective German customer to build an information retrieval system based on Metal technology. Another intriguing direction is that of Controlled English applications. The Metal group in Liège (Belgium), which developed the Dutch-French system at the behest of the Belgian government, is prime contractor in an EU LRE project called SECC of which the goal is to develop a controlled English checker based on Metal, a tacit acknowledgement that Sietec has one of the best English parsers in the industrial world. It may seem odd that the Commission is funding the kind of development work which has already been done by other groups in industry but it certainly is one way to be assured of getting results. However, as always with the Commission's cost-shared funding efforts, it is never exactly clear who will be allowed to exploit the fruits of such an endeavour.

Large customers, however, are comparatively few and far between, so Sietec is also looking at ways of deploying its technology in more modest arenas. Off-the-shelf packages are, after all, the bread-and-butter (and the future) of the software industry. Here Sietec has less experience. Four years ago, it launched a terminology management package for the PC, called Term-PC. More recently, it introduced LingTools, a Unix-based client/server package. LingTools is a set of four modules for manipulating and managing terminology: LexikonAbfrage, a look-up tool which enables users to develop their own glossaries in conjunction with standard lexicons; LexikonTextVergleich, a tool for comparing a text against a lexicon; IndexGenerierung for automatically indexing a text; and LexikonAdministration, a tool for creating and maintaining terminology lists. At the moment, only a Unix/Motif version of LingTools is available, but a Windows client is currently being beta-tested and a Windows server is under development.

Sietec has been extensively involved in the Eurolang project, to which it has dedicated some twenty engineers, thereby making it Site's major industrial partner in the undertaking. Initially -and up until quite recently - the consortium spoke of building a "Eurolang MT System", with Metal as the translation engine as the system's core. This would

have been the upgrade for both current Metal users and new customers. However, Sietec and Site have tempered their collective ambitions somewhat. Sietec will continue to develop Metal in its current form and offer the system as an optional back-end to Optimizer, while Site will offer Optimizer as a front-end to Metal and other MT systems. Whichever way the cake is sliced, Optimizer remains of strategic importance to Sietec.

Further developments of Metal will continue to be a major focus for Sietec. New language pairs are under way, including Catalan at Sietec's former subsidiary in Barcelona, now called Incyta; Arabic is another possibility. Many of the new language pairs are being developed in collaboration with third parties; for German-Danish, for example, is now under development at the Institut for Erhvervsforskning, in Koling (Denmark).

With its huge investment in language processing, Sietec is one of a small handful of companies with both the experience, the technology, and the internal resources to exploit the burgeoning NLP market. Its ultimate survival will largely depend on how quickly it can deploy robust high-end systems and simultaneously downsize and "package" its technology for the much larger small-system market.

TOPTRAN: a custom-built system

[Based on article in AAMT Journal no.7, June 1994]

The Toppan Printing Company became dissatisfied with a proprietory English-Japanese system, finding it "inconvenient to use and difficult to handle", so they decided to develop its own system - the "User-Oriented Machine Translation System TOPTRAN." A major problem with the commercial MT system was that improvements could only be made by the makers, and they were unwilling to respond to all requests for changes. Toppan was also dissatisfied by the amount of pre-editing and post-editing necessary to produce output of the quality required by a printing firm.

The facilities sought by their translators were that:

1. Translation quality should be good enough for a rough draft. The system should not aim for perfection, as even human translations are edited. Perfectionism leads to excessive complexity in MT algorithms.

2. Ordinary translators should be able to make improvements without needing to know much about the system.

3. The system should be able to learn from corections made during post-editing. "Once a sentence has been properly translated the same translation should be output for the same sentence next time it appears."

4. Pre-editing should not involve work that is not normally done in manual translation.

5. System operability should be equivalent to that of a word processor.

Toppan adopted an example-based MT approach, because: (i) "translations are performed by finding a very similar translation example and imitating it", (ii) "translation ability is increased simply by adding new translation examples", and (iii) "through imitation of human-translated sentences, natural translations are output." Toppan's EBMT type system called TOPTRAN (Toppan Translation Database System) is designed for English-Japanese and Japanese-English translation, and has currently the following features. "Bilingual correspondence recording" involves the aligning of parallel English and Japanese translations, allowing for exact matching. In "bilingual correspondence learning", the aligned sentences are automatically broken down into potential corresponding words and phrases (not sysntactic units), and the correspondences are then confirmed or not by operators as correct

equivalents. The "translation" function is based on the biligual database; there is no pre-editing; the closest bilingual correspondence data is searched for and retrieved on a score of similarity (there is no use of a thesaurus.)

The system has been evaluated on a NSSUN-SP2, with results indicating that about 75% of the output was considered usable as rough draft translations. At the present time there are several thousand bilingual correspondence pairs; and speed of production is claimed to be significantly faster than commercial MT systems on the market, although the effect of a larger database is unknown.

RECENT US PATENTS, 1986-1994

The list which follows is intended to be the first of a regular feature in MT News International. In this listing we give the titles of the patents, the US patent numbers, the dates of patents, the names of inventor(s) and assignee(s), the dates of filing, and brief extracts from the patents. In each case, extracts are taken from the `Abstract', the `Background of the invention' and/or the `Summary of the invention', and they are intended to give readers some general idea of the content of the patents and the claims. They cannot be complete in any sense, and readers interested in specific patents should consult the originals. For information about the nature and legal force of US patents, which differ substantially from those in Europe and Japan, we refer you to the article by Leighton K.Chong in MTNI#7: 14-15, and to his article on "Intellectual property rights and MT" given at the AMTA-94 conference in October 1994.

Patent number: 4,599,691
Date of patent: Jul.8, 1986
TREE TRANSFORMATION SYSTEM IN MACHINE TRANSLATION SYSTEM
Inventors: Hiroshi Sakaki, Kazuo Hashimoto
Assignee: Kokusai Denshin Denwa Co., Ltd., Tokyo, Japan
Filed: May 13, 1983
1 claim, 135 drawing figures

... As for machine translation system, many systems have been proposed so far. The most popular ones have a feature that they generate target language ... after introducing sentence in source language... and applying the process consisting of extracting the meaning of the introduced sentence in source language, which process should be called processing into deep structure. Though these types of methods conform with theories in linguistics, they are not well suited to accept special expressions a little out of general rules of language in spite of the fact that these expressions inevitably occur in natural language such as Japanese, and to reflect these expressions to translation.

... This invention treats language as a set of specific patterns and conducts translation by allotting corresponding patterns in target language for each pattern in source language, and as a result, can conduct sophisticated translation by including patterns corresponding to these specific expressions.

Patent number: 4,661,924
Date of patent: Apr.28, 1987
MULTIPLE-PARTS-OF-SPEECH DISAMBIGUATING METHOD AND APPARATUS FOR MACHINE TRANSLATION SYSTEM
Inventors: Eri Okamoto, Atsushi Okajima
Assignee: Hitachi, Ltd., Tokyo, Japan
Filed: Jul.19, 1985
15 claims, 12 drawing figures

... According to the present invention, the part-of-speech disambiguation and hence the syntax analysis or parsing of sentences written in natural language can be accomplished with a high accuracy by virtue of such an inventive feature that the parts of speech the words should be in a given sentence are determined on the basis of the frequency at which the parts of speech makes appearance for the words while conforming to the grammatical rules of the language in which the sentence is written. Further, in case a number of sentences belonging to a same sphere of literatures written in a natural language are to be parsed, the frequencies at which the parts of speech make appearance in the sentences are updated on the basis of the result obtained from the preceding part-of-speech disambiguation processing and syntax analysis of the sentences belonging to the same or similar sphere of literatures can be performed with a corresponding increased accuracy.

Patent number: 4,706,212
Date of patent: Nov.10, 1987
METHOD USING A PROGRAMMED DIGITAL COMPUTER SYSTEM FOR TRANSLATION BETWEEN NATURAL LANGUAGES
Inventor: Peter P.Toma
Filed: Mar.31, 1971
24 claims, 37 drawing figures

... The present invention, hereinafter called SYSTRAN was possible because of the combination of experiences as systems analyst, linguist and computer programmer... The unique solution which is elaborated upon herein can be summarized as follows: computerized syntax, a dictionary lookup which is based on the most efficient exploitation of main and auxiliary storages and which uses four different types of dictionaries, the combination of fixed and variable length areas for full utilization of the computer, the introduction of remembering switches, new ways of handling multiple meaning problems, and system translation (SYSTRAN), universality which makes it possible to change only tables and parameters in most instances when source and/or target languages are changed.

Patent number: 4,774,666
Date of patent: Sep.27, 1988
TRANSLATING APPARATUS
Inventors: Kouiji Miyao, Hitoshi Suzuki, Hazime Asano, Shinji Tokunaga, Yauhiro Takiguchi, Shuzo Kugimiya
Assignee: Sharp Kabushiki Kaisha, Osaka, Japan
Filed: May 14, 1986
1 claim, 13 drawing sheets

A translating apparatus comprising a synonym display device for displaying synonyms related to a certain translated term in a translated sentence, and a replacing arrangement for replacing the certain translated term with an appropriate synonym selected among the synonyms displayed by the display device, whereby a more suitable translated term or terms can be selected through consultation of a dictionary contained in the apparatus.

Patent number: 4,787,038
Date of patent: Nov.22, 1988
MACHINE TRANSLATION SYSTEM
Inventors: Miwako Doi, Tsutomu Kawada
Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan
Filed; Mar.24, 1986
7 claims, 5 drawing sheets

In a machine translation system, a translated-sentence possibility corresponding to an input original sentence is obtained by a translation processing section with reference to a dictionary, and is displayed to be divided by a division display control section into processing units in the translation processing section. When some portions of the original sentence have other translated-word possibilities in addition to the displayed translated-word possibilities, display control is performed by

an identifying display control section so that the portions having other translated-word possibilities can be distinguished from portions having no such other translated-word possibilities and that classes of portions having the other translated-word possibilities are identified. Classes of the other translated-word possibilities are displayed in accordance with a predetermined priority order.

Patent number: 4,805,132
Date of patent: Feb.14, 1989
MACHINE TRANSLATION SYSTEM
Inventors: Toshio Okamoto, Kimihito Takeda
Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan
Filed: Aug.19, 1986
6 claims, 4 drawing sheets

In a machine translation system of the present invention, an original sentence input from an input section is stored in an original storage section and is subjected to translation processing using linguistical information stored in a dictionary section. A translated sentence obtained through the translation processing is stored in a translation storage section. Then the original and translated sentences are respectively displayed on predetermined display regions of a display section, and are subjected to edit processing in accordance with information supplied from the input section. In addition, the machine translation system detects an information input position of the information input from the input section on the display section, and discriminates from the input position whether the input information is a newly input original sentence. If the information is a newly input original sentence, the original sentence is stored in the original storage section. If it is discriminated that the information is not a newly input original sentence, translation/edit processing corresponding to the input information is performed... Therefore, a switching operation for operation modes of the system can be omitted, and an original sentence can be input even during translated-sentence edit processing. In addition, inputting of an original sentence and instruction for translation/edit processing can be selectively used upon control of an input position, i.e. a cursor position on the display section.

Patent number: 4,814
Date of patent: Mar.21, 1989
MACHINE TRANSLATION SYSTEM TRANSLATING ALL OR A SELECTED PORTION OF AN INPUT SENTENCE
Inventors: Shinobu Shiotani, Yoji Fukumochi, Shinji Tokunaga, Syuzo Kugiyama, Ichiko Sata, Hitoshi Suzuki
Assignee: Sharp Kabushiki Kaisha, Osaka, Japan
Filed: May 20, 1987
9 claims, 3 drawing sheets

A machine translation system to display an input sentence and a translated sentence thereof, comprising a region specifying circuit for specifying a desired portion in the input sentence, and a partial translating circuit for translating only the portion specified by the region specifying circuit is disclosed.

... This invention relates to a machine translation system capable of promoting translation works in an interactive method between the operator and the system, by displaying an input sentence (original) and the result of translation of this input sentence, that is, the translated sentence on the CRT screen. In particular, the present invention permits a translating region of the input sentence to be arbitrarily specified.

Patent number: 4,821,230
Date of patent: Apr.11, 1989
MACHINE TRANSLATION SYSTEM
Inventors: Akira Kumano, Hiroyasu Nogami, Seiji Miike, Shin-ya Amano
Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan
Filed: Jan.2, 1987
10 claims, 9 drawing sheets

It is an object of the present invention to provide a machine translation system which can appropriately provide a translated sentence with punctuation marks (regardless of the presence, absence, or positions of punctuation marks in the original sentence), and can conveniently obtain a natural-sounding translated sentence which is easy to read.

Patent number: 4,831,529
Date of patent: May 16, 1989
MACHINE TRANSLATION SYSTEM
Inventors: Seiji Miike, Noriko Yamanaka, Hideki Hirakawa, Shinya Amano
Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan
Filed: Feb.12, 1987
9 claims, 5 drawing sheets

The present invention relates to a machine translation system for performing correct and speedy translation by the use of a plurality of different types of dictionaries for translation.

It is... an object of the present invention to provide an improved machine translation system in which translated sentences in good quality can be obtained rapidly by the effective use of information contained in both the general purpose and nomenclature dictionaries.

It is another object... to provide an improved machine translation system in which the priority order of at least two kinds of dictionaries to be used is determined in accordance with the kinds of parts of speech and the contents of information used in the original as well as the conditions of the two dictionaries.

It is still another object... to provide an improved machine translation system in which a suitable translation can be obtained promptly without necessitating the post edition of the translation, otherwise necessary, by simplifying the edit operation.

Patent number: 4,833,611
Date of patent: May 23, 1989
MACHINE TRANSLATION SYSTEM
Inventors: Yoji Fukumochi, Shinji Tokunaga, Hitoshi Suzuki, Syuzo Kugimiya, Shinobu Shiotani, Ichiko Sata
Assignee: Sharp Kabushiki Kaisha, Osaka, Japan
Filed: Jun.26, 1987
8 claims, 5 drawing sheets

A machine translation system for translating inputted sentences using a translation dictionary, defined grammatical rules and tree conversion rules. The translation system is capable of producing partial translation when the entire sentence cannot be translated as a whole. Each translated portion is representative of an analyzed partial tree of the complete inputted sentence.

Patent number: 4,864,503
Date of patent: Sep.5, 1989
METHOD OF USING A CREATED INTERNATIONAL LANGUAGE AS AN INTERMEDIATE PATHWAY IN TRANSLATION BETWEEN TWO NATIONAL LANGUAGES
Inventor: Bruce G. Tolin
Assignee: Toltran Ltd., Barrington, Ill.
Filed: Feb.5, 1987
29 claims, 3 drawing sheets
Abstract:

A method of using a created international language as an intermediate pathway in translation between two national languages, wherein the first language is translated into the created international language, the latter serving as a bridge, which is then translated into the second language. In the present method there are interpretive steps rather than a strict word for word translation, without interpretation, which has proven fallacious in the past.

... Such a translation is reversible in either direction and can accommodate translation from one national language into the created international language and then translate into a multiplicity of second national languages from the created international language.

Patent number: 4,964,044
Date of patent: Oct.16, 1990
MACHINE TRANSLATION SYSTEM INCLUDING SEMANTIC INFORMATION INDICATIVE OF PLURAL AND SINGULAR TERMS
Inventors: Akira Kumano, Yumiko Sugiura, Chiaki Aoyama
Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan
Filed: Sep.22, 1989
8 claims, 7 drawing sheets

To translate Japanese having no morphological distinctions between singular and plural forms into English having a morphological distinction between the two, for instance, the dictionary unit includes semantic information indicative of the plural number, and the translation unit syntactically and semantically translates Japanese into English as follows: a Japanese sentence is morphologically analyzed into basic morphemes by an inflection information dictionary; lexical syntactic information of the Japanese basic morphemes are retrieved from a word/phrase dictionary; the Japanese sentence is syntactically analyzed into an intermediate structure by Japanese syntactic analysis grammar to clarify modification relationship between tow words; the Japanese intermediate structure is semantically transferred into an English intermediate structure by structure transfer grammar; a concept data of a noun (e.g. "book") which includes FEATURE: NUMBER=PLURAL is formed when the noun is modified by an adjective (e.g." many") indicative of the plural number; an English sentence structure is syntactically generated from the intermediate structure and by English syntactic generation grammar; and lastly English morphemes are generated by morphological generation grammar to change "book" into "books".

Patent number: 5,088,038
Date of patent: Feb.11, 1992
MACHINE TRANSLATION SYSTEM AND METHOD OF MACHINE TRANSLATION
Inventors: Katumi Tanaka, Hideki Hirakawa, Hiroyasu Nogami, Shin-ya Amano
Assignee: Kabushiki Kaishan Toshiba, Kawasaki, Japan
Filed: May 24, 1990
10 claims, 9 drawing sheets

A machine translation system capable of obtaining a consistent translation for an entire document by taking context into account in translating each word or sentence. In this system, document information necessary to remove uncertainty in the translation due to a presence of a plurality of candidates for the translation is utilized whenever uncertainty due to a presence of a plurality of candidates for the translation arises by attempting to translate according to a translation dictionary containing rules for translation.

Patent number: 5,136,503
Date of patent: Aug.4, 1992
MACHINE TRANSLATION SYSTEM
Inventors: Akira Takagi, Jun Takada, Minoru Yukawa
Assignee: Kabushiki Kaisha CSK, Tokyo, Japan
Filed: Dec.19, 1989
6 claims, 5 drawing sheets

A machine translation system for automatically effecting translation from Japanese the source language into another or target language. Conventional translation systems entail problems, e.g. the reduction in processing efficiency owing to the need for an additional step such as pre-editing prior to translation. In the machine translation system of the present invention, the main process step for mechanical translation, the syntactic analysis step, is based on the determination of whether or not any modification relationship can be established for each pair of adjacent words by extracting a rule

corresponding to that combination from a part of speech matrix table in which analysis rules of the source language are described with respect to the parts of speech of the corresponding pair of words to modifying words and head words; successively stacking each of these word combinations for which a modification relationship has been established for use as an analysis tree element; and displaying an analysis tree corresponding to the original sentence in the source language from the stacked analysis tree segments. The number of syntactic rules is thereby limited to the square of the number of individual parts of speech, the need for pre-editing is eliminated and Japanese sentences in every style can be translated with improved efficiency.

Patent number: 5,140,522
Date of patent: Aug.18, 1992
METHOD AND APPARATUS FOR MACHINE TRANSLATION UTILIZING PREVIOUSLY TRANSLATED DOCUMENTS
Inventors: Etsuo Ito, Koichi Hasebe, Shinya Amano
Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan
Filed: Oct.27, 1989
14 claims, 21 drawing sheets
Abstract:

A method and apparatus for machine translation of a language capable of dealing efficiently with situations which conventionally required repetitions of similar translations...

... there is provided a method of machine translation, comprising the steps of: (a) translating each original document is a first language into a corresponding translated document in a second language; (b) storing all the original documents and the translated documents in corresponding arrangements; (c) for each new original document to be translated, searching the original documents stored at step (b) which are utilizable in translation of the new original document to be translated; and (d) translating the new original document to be translated by utilizing the stored translated documents corresponding to the stored original document found at the step (c).

Patent number: 5,175,684
Date of patent: Dec.29, 1992
AUTOMATIC TEXT TRANSLATION AND ROUTING SYSTEM
Inventor: Leighton K.Chong
Assignee: Trans-Link International Corp., Honolulu, Hi.
Filed: Dec.31, 1990
23 claims, 3 drawing sheets

A machine translation system includes an automatic routing system for automatically translating input text of a source language and sending output text of a selected target language to an addressee, in response to designations of the addressee, the target language, and/or the selected sublanguage scanned from a cover page for the input text. The system can interface with a number of input sources, such as a fax machine combined with a page scanner, a fax/modem board, or a network interface coupled to a server computer system. The server computer system can control the operation of the machine translation module for multiple target languages and/or sublanguages as designated by the cover page. A page formatting module is used to compose the translated text into a desired page format, and to provide footnotes for alternate translations of ambiguous phrases of the input text. The cover page can also designate a plurality of addressees in different target languages. The output end of the system may be a fax/modem board for sending output text to an addressee's fax number, a network interface, or a page printer for printed copy. The machine translation module(s) of the system may be upgraded or replaced without disruption to the overall telecommunications routing functions of the system.

Patent number: 5,214,583
Date of patent: May 25, 1993
MACHINE LANGUAGE TRANSLATION SYSTEM WHICH PRODUCES CONSISTENT TRANSLATED WORDS

Inventors: Seiji Miike, Hideki Hirakawa, Etsuo Ito, Sin-ya Amano
Assignee: Kabushiki Kaisha Toshiba, Kawasaki, Japan
Filed: Nov.22, 1989
19 claims, 11 drawing sheets

A machine translation system including an input unit, a first text memory, a translation dictionary, a translation processor, a second text memory, and a control editor... The control editor includes a changing section... and a replacing section... When a translated word corresponding to a certain word or phrase in an original text is changed, the changed portion of the translated word and a corresponding original word or phrase are extracted, and entry words including the extracted original word or phrase in a word/idiom dictionary are retrieved. The translated word portions corresponding to the original word or phrase in the retrieved entry words are replaced with the extracted translated word portion.... As a result, when a translation is performed using the word/idiom dictionary, a translated text in which translated words corresponding to the same word or phrase are consistent can be obtained. Therefore, a correct translated text which is easy to read can be easily obtained. Since a translated text with consistent translated words can be obtained by only changing a translated word, translation efficiency can be greatly improved as compared to an operation wherein translated words are corrected for consistency while observing a translated text after a translation result is obtained.

Patent number: 5,224,039
Date of patent: Jun.29, 1993
METHOD OF ENABLING THE TRANSLATION OF MACHINE READABLE INFORMATION
Inventors: Daryl R.Cox, Francis M.Fandrick
Assignee: International Business Machines Corporation, Armonk, N.Y.
Filed: Dec.28, 1990
5 claims, 2 drawing sheets

... To compete in the increasingly important global markets, it is necessary to be able to transform software written in the United States for use in other countries. One of the clearest differences between versions of the same product in various countries is seen in the text displayed on the computer display screen. This text, which includes messages, prompts, user keyboard input, help information, menus, and the like, is called machine readable information....

In the present invention, U.S. developers can develop their machine readable information in English and the proper expansion space is inserted as appropriate for each language. More specifically,... a table of rules is formed that sets forth the number of additional spaces required to contain the second national language translation for a phrase of machine readable information written in the first national language. Then, an appropriate number of additional spaces taken from the table is automatically added to each phrase of the machine readable information. Thus, the U.S. developer can design his or her screen without regard to translation, saving him or her time and effort. Then, the method of the present invention expands the machine readable information file and the developer can resize the screen appropriately.

DATABASES and SERVICES

Intergraph MT System Available on CompuServe

[Press release from CompuServe]

(Columbus, Ohio, Sept. 21, 1994) -- In an industry first, CompuServe, the only global commercial information service, today announced the availability of multi-lingual machine translation in its MacCIM Help Forum. Translation from English to French and French to English, as well as English to German and German to English, is done automatically in near

real-time, making it possible for CompuServe users who do not speak the same language to share information and understanding.

"CompuServe is a global community, bringing people and ideas together across time zones and cultures," said Timothy Oren, CompuServe's vice president, Future Technologies. "This new technology is a first step toward removing the language barriers to worldwide communication and understanding. It demonstrates our ongoing commitment to serving a global community."

"It's important to realize that this represents the very early stages of a new technology," cautioned Mary Flanagan, Ph.D., CompuServe's Machine Translation project leader. "Translation by computer is difficult because human language is often complex and ambiguous. Many words and phrases have different meanings depending on their context.

"In addition, long and complex sentences are very difficult for computers to analyze," Flanagan added. "As a result, the quality of translations done by computer is much lower than human translations. The key benefit is the timeliness of machine translation."

Forums are online discussion areas within CompuServe, where members converse, exchange questions and offer advice. Typical forum conversations are completed within 24 hours, and in the case of software support forums, the timeliness of replies is crucial. The MacCIM Forum is a part of CompuServe's basic services, and offers support and assistance for users of the CompuServe Information Manager for the Macintosh (MacCIM).

"The value of exchanges in forums is their immediacy," Flanagan said. "Human translation takes time, and in the context of a support forum, messages lose relevance and value if they are delayed. The purpose of machine translation is to quickly provide usable quality text that helps a non-English speaker understand messages they would otherwise have no way of reading."

For its online translation service, CompuServe uses a commercially available natural language processing technology from Intergraph Corporation. "The Intergraph system was selected after an extensive evaluation of machine translation systems," Flanagan said. "We liked its output quality, translation speed and adaptability."

Messages to be translated are picked up from the forum hourly and translated at a rate of up to 1,800 words per minute. Turnaround time depends on the number of messages to be translated, but currently is only a few minutes. French and German speakers can specify their preferred language in their user profile, and they will be automatically routed to the version of the MacCIM forum in their own language. Because of the imperfect nature of machine translation, the writer's original message is provided along with the translation.

"It helps the quality of the translation if the writer uses clear, simple sentences and unambiguous words," Flanagan explained. "If the writer is attuned to these issues, the quality of the translation is improved."

Machine translation services are presently available in the MacCIM Help Forum, and CompuServe will be extending the technology to its electronic mail and File Finder services, as well as expansion to other languages in the future. "The languages we chose to begin with reflect our user population," Flanagan said. "These languages represent a significant non-English speaking member base."

Intergraph, a Fortune 500 company, develops and sells computer systems that range from personal desktop tools to full-scale integrated solutions for publishing, engineering and other technical applications.

CompuServe members pay $8.95 per month for unlimited connect-time access to 70 basic services. Members can access CompuServe's other 2,000 extended services for $4.80/hour with a 1200 or 2400 bits per second modem and for $9.60/hour with a 9.6 or 14.4 kilobits per second modem. Established in 1979, the CompuServe Information Service

provides its worldwide membership of 2.2 million with databases and services to meet both business and personal interests. CompuServe can be accessed by any modem-equipped personal computer using general communications software. In addition to the CompuServe Information Service, CompuServe Incorporated provides frame relay, wide area networking services, electronic mail, entertainment and business information services to consumers and major corporations worldwide. CompuServe is an H&R Block (NYSE:HRB) company.

Contact: R. Pierce Reid or Michelle Moran, CompuServe Incorporated, 614/538-4571 or 614/538-3497

[Press release from Intergraph]

Machine Translation (MT) Technology Gaining Acceptance in On-line World. (Huntsville, AL, December 1, 1994) -- Intergraph Corporation today announced the growing acceptance of the on-line machine translation (MT) service of CompuServe's MacCIM Support Forum. After an initial phase of challenging the translation software with literature excerpts and the like, users are applying the service as a business tool to translate help messages and other information into their native language. The first on-line forum to offer translation, the MacCIM Support Forum began offering the translation service on August 23. To date, more than 15,000 messages have been translated using the on-line service.

The MacCIM Support Forum uses Intergraph's translation-enabling technology for bi-directional translation of messages between English and French and between English and German. Messages to be translated are picked up from the Forum hourly and translated with Intergraph's MT software at a rate of up to 1,800 words per minute. In addition, the MacCIM Support Forum offers support and assistance for users of the CompuServe Information Manager for the Macintosh in multiple languages.

Mary Flanagan, Ph.D., CompuServe's Machine Translation project leader, said the use of the Forum's translation service has evolved from a novelty to a 'working' tool. "At first, users were just experimenting with the translation software, submitting poems, idiomatic expressions, song lyrics, tongue twisters and the like. Lately," she continued, "we've seen a shift to more serious uses of the service. This trend demonstrates that MT can be used productively in an on-line environment and shows its potential as a medium for exchanging information among multilingual users."

According to Flanagan, French- and German-speaking users have begun asking support questions in their native language. The messages can then be read and processed by the Forum's English-speaking System Operator, who manages and monitors Forum contents. The software translates the messages to English, and the System Operator replies. The software then translates the System Operator's reply to French or German so users can read the response in their own language. In another use of the translation service, some users have begun translating documents by posting the text as a message to themselves and then collecting the results.

CompuServe's Forum translation service is an example of the expanding role of Intergraph's translation-enabling technology. Deane Dayton, Ph.D., manager of Intergraph's language translation products, said, "In the past few years, Intergraph has been developing its expertise in language processing and is supporting its MT product at organizations such as McDonald's and AT&T. In the future, we see MT moving more and more into the business mainstream, and applications like CompuServe's on-line service help users see the potential of this technology in a new way."

A member of the Fortune 500, Intergraph Corporation (Huntsville, Ala.) is the world's largest company dedicated to supplying interactive computer graphics systems. Products

range from point solutions, meeting individual and departmental needs, to integrated, enterprise-wide systems.

Noted for delivering interoperable systems and applications, Intergraph bases its products on Windows, Windows NT, and UNIX operating systems.

One of Intergraph's four business units, Intergraph Software Solutions (ISS) develops and markets integrated software for the Technical Desktop - the combination of compatible technical applications and personal productivity tools in a single desktop computer. Technical applications include computer-aided design, engineering, analysis, manufacturing, publishing and earth sciences. ISS also provides core system software, high-end applications, and training, consulting, and implementation services.

CompuServe Background Information

Established in 1979, the CompuServe Information Service provides its worldwide membership of 2.4 million with databases and services to meet both business and personal interests. CompuServe can be accessed by any modem-equipped personal computer using general communications software. In addition to the CompuServe Information Service, CompuServe Incorporated provides frame relay, wide area networking services, electronic mail, entertainment and business information services to consumers and major corporations worldwide.

Intergraph is a registered trademark of Intergraph Corporation. Other brands and product names are trademarks of their respective owners.

For further information: Susan Moore (Intergraph, (205) 730-3315); Pierce Reid (CompuServe, (614) 457-8600)

Globalink on CompuServe

[From *CompuServe Magazine* October 1994]

The Internet allows its users to exchange e-mail and gain access to a vast wealth of online information. The information superhighway now also allows users to translate their documents between several languages.

Globalink, a Fairfax, Virginia-based company, has launched the Foreign Language Message Translation System (MTS). The service allows registered users to send messages and text files on the Internet and rapidly receive draft translations of their documents from English into Spanish, French, or German, or vice versa.

To access MTS, users register by fax or e-mail using a credit card. MTS is available on a 24-hour basis. The service costs five cents per word, with a minimum charge of $5. For added translation quality, users can request that a translation be checked against an industry-specific dictionary. MTS currently offers 18 dictionaries, including Legal, Computer, Medical, Automotive, and Business/Finance.

Globalink says its translation software is based on sophisticated algorithms that can detect a word's meaning based on its context in a full sentence. The code can also handle gender, pluralization, verb conjugation, and other grammatical structures.

"It's a convenient, cost-effective tool for users who need fast turnaround on language translations," observes Larry Golfer, director of business development. "While machine-assisted translations cannot be expected to provide 100 percent accurate results, the MTS service will provide clear, understandable, and useful draft translations."

More details are available from Globalink via the Internet. Write to info@glnk.com.

British National Corpus

[From CORPORA list]

Several people have asked about the current status of the British National Corpus, following some rather inaccurate stories about it in the British press. For those who may have missed even the accurate stories, the BNC is a very large (100 million words) corpus of modern English, both spoken and written, produced by an academic/industrial consortium led by Oxford University Press, involving Longman UK Ltd, Chambers/Larousse, Oxford University Computing Services, the University of Lancaster and the British Library. Production of the corpus was funded by the commercial partners and by the UK Government, under the DTI/SERC Joint Framework for Information Technology.

Design. The corpus is intended to represent as many of the varieties of modern English as possible. The written part (90%) includes for example, both local and national newspapers, academic books and popular fiction, published and unpublished letters, memoranda, school and university essays etc etc. The spoken part (10%) includes unscripted informal conversations as well as formal business meetings or lectures.

Tagging. At the last count, the corpus contained 104 million words, totalling about 1.6 gigabytes of disk space. The corpus is automatically segmented into orthographic sentence units, and each word in the corpus is automatically assigned a word class (part of speech) code by the CLAWS software developed at the University of Lancaster. The corpus is encoded according to the TEI (Text Encoding Initiative)'s Guidelines, using the ISO standard SGML to represent this and a variety of other structural properties of texts (e.g. headings, paragraphs, lists etc.). Full classification, contextual and bibliographic information is also included with each text in the form of a TEI conformant header file.

Availability. The corpus will be freely available for academic research purposes within the European Union, probably by the end of this year. We will be distributing both the whole corpus (on tape) and a 1% sampler (on CD). In both cases, a specially-written software package (SARA) will be provided as well, but the BNC can be processed using any SGML-aware software system, as well as by standard Unix or DOS utilities.

The text of the corpus is now complete. We are currently trying to supply bibliographic and contextual information information which is missing for several texts. We have an alpha version of the SARA access software running, which has revealed some problems in the software and indexing process which need to be fixed before releasing it.

If you would like to be notified as soon as the BNC is available for release, please send your name and contact information to: British National Corpus, Oxford University Computing Services, 13 Banbury Road, Oxford OX2 6NN, UK. (Tel: +44 (1865) 273 280; Fax: +44 (1865) 273 275; Email: natcorp@oucs.ox.ac.uk

---

English Corpora: a summary list

Kathy Mitchell

[The following useful summary is extracted from the LINGUIST list, 5-1186]

A 1993 survey of well-known corpora was written by Jane Edwards and published as chapter 10 in the book: Edwards, Jane A. & Martin D. Lampert (eds). TALKING DATA: TRANSCRIPTION AND CODING IN DISCOURSE RESEARCH. London and Hillsdale, NJ: Erlbaum. 336 pp. ISBN: 0-8058-0349-1. This chapter is also available via anonymous ftp from cogsci.berkeley.edu in compressed format, in the "pub" directory, under the filename of "CorpusSurvey.Z and as the LINGUIST file "CORPORA FAQ", which is retrievable by email by sending the message "GET CORPORA FAQ LINGUIST" to LISTSERV@tamvm1.

tamu.edu (For a list of all the archived LINGUIST files, send the message "INDEX LINGUIST" to this same address.)

A couple other surveys of corpora were mentioned as well. The Lancaster Survey of Machine-Readable Language Corpora, is available from the ICAME file server FAFSRV@NOBERGEN.BITNET (send mail with Subject: DIR)

Another survey "to create and maintain a comprehensible database about archives and projects in machine-readable text" was done by the Center for Text and Technology at Georgetown University, in collaboration with other centres. More information can be obtained at: Michael Neuman, Ph.D., Georgetown Centre for Text and Technology, Reiss Science Building, Room 238, Georgetown University, Washington, DC 20057, U.S.A.

There is an unmoderated email list, CORPORA, for discussion about text corpora such as availability, aspects of compiling and using corpora, software, tagging, parsing, bibliography, etc. One can subscribe by sending an email message with the command "sub corpora <firstname> <lastname>" to LISTSERV@UIB.NO This list is hosted at the Norwegian Computing Centre for the Humanities in Bergen, Norway. Information stored at the machine nora.hd.uib.no can be accessed through gopher, anonymous FTP, or a mail server (send a "help" message for more details). There is also a World-Wide-Web page with the URL http://www.hd.uib.no A contact for any of these services is Knut Hofland (knut.hofland@.hd.uib.no)

The International Computer Archive of Modern English (ICAME) has a number of English corpora (including American, British and Indian English corpora) available in various media for a low cost, primarily for research and teaching purposes. It is distributed by the Norwegian Computing Centre for the Humanities (NCCH) in Bergen, Norway, which can be contacted at icame@hd.uib.no

The Linguistic Data Consortium has an extensive list of corpora available for sale. Info about these can be gotten by anonymous ftp from ftp.cis.upenn.edu in directory /pub/ldc, or by email from ldc@unagi.cis.upenn.edu.

The Center for Electronic Texts in the Humanities (CETH), a joint Rutgers/Princeton organization is worth investigating. They can be contacted at ceth@pucc.princeton.edu or hockey@zodiac.rutgers.edu.

There is some interesting tagging software available by anonymous ftp from PARCFTP.Xerox.COM in /ftp/pub/tagger.

Some specific corpora mentioned were:

The Multilingual Corpus 1 of the European Corpus Initiative (ECI/MCI) contains almost 100 million words in 27 (mainly European) languages. It consists of 48 opportunistically collected component corpora marked up in SGML. The CD-ROM is available in the US from the Linguistic Data Consortium (LDC) or from ELSNET, 2 Buccleuch Place, Edinburgh EH8 9LW, SCOTLAND. Information on ordering it from ELSNET can be obtained from elsnet@cogsci.ed.ac.uk; or http://www.cogsci.ed.ac.uk/elsnet/eci.html; or by anonymous ftp from ftp.cogsci.ed.ac.uk:pub/elsnet/eci/mci-listing

The ACL has a CD-ROM, in ISO 9660 format, containing about 300 Mb of Wall Street Journal text, a large collection of scientific abstracts, the full text of the 1979 edition of the Collins English Dictionary, and some samples of tagged and parsed text from the Penn Treebank project. To order this, send a message to Rafi Khan (khanr@unagi.cis.upenn.edu) including your mailing address, and he will send a paper copy of the User Agreement to be signed.

The SUSANNE Corpus is an annotated sample comprising about 130,000 words of written American English text, produced to exemplify a set of annotation standards which attempt to specify an explicit notation for all aspects of the surface and logical grammar of

real-life English in sufficient detail that analysts independently applying the standards to the same text must produce identical annotations. These standards are defined in the book ENGLISH FOR THE COMPUTER; a skeleton outline of the scheme is included in the electronic documentation file which accompanies the Corpus. The texts of the SUSANNE Corpus are a subset of the texts included in the (unannotated) Brown University Corpus. Release 3 of the SUSANNE Corpus is available is available by anonymous ftp from the Oxford Text Archive at black.ox.ac.uk in the directory ota/susanne - follow the instructions in the README file in that directory.

The Oxford Text Archive is a repository for some corpora as well: Oxford Text Archive, Oxford University Computing Service, 13 Banbury Road, Oxford OX2 6NN. E-mail: ARCHIVE@VAX.OX.AC.UK

Dutch Corpus Available from INL

[From: LINGUIST List, 5-1379]
The Institute for Dutch Lexicology INL offers you the possibility to consult a text corpus of ca. 5 million words of present-day Dutch text, by the international computer network. This corpus is different from the Dutch INL corpora on the ECI/MCI CD-ROM distributed by the Linguistic Data Consortium and ELSNET. The texts are derived from books, magazines, newspapers and TV broadcasts, and cover several topics such as journalism, politics, environment, linguistics, leisure and business & employment. You can easily define subcorpora on the basis of these parameters.

The retrieval system allows you to search for single words or for word patterns, including some predefined syntactic patterns that can be changed by the user. Searches concern the levels of word form, part of speech (POS), and head word, both separately and in combination by use of Boolean operators and proximity searches. During the search, data concerning frequency and distribution over the texts are provided at several levels. The output most often is a list of items, or a series of concordances (words in context) with a variable, user-defined textual context. Sorting facilities may make your analysis of the output data more comfortable. 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.

The providers of the texts have given permission for use of their materials for non-commercial, research purposes only. The conditions for commercial use are still topic of discussion.

In order to get access to this corpus, an individual user agreement has to be signed. An electronic user agreement form can be obtained from our mailserver Mailserv@Rulxho.Leidenuniv.NL. Type in the body of your e-mail message: SEND [5MLN94] AGREEMNT.USE. For Dutch users a Dutch version is available in the same directory. The filename is OVERKMST.GEB. Please make a hard copy of the agreement form, sign it, and return the signed copy to: Institute for Dutch Lexicology INL P.O. Box 9515, 2300 RA Leiden fax: 31 71 27 2115. After receipt of the signed user agreement, you will be informed about your username and password. If you need additional information, please send an e-mail message to Helpdesk5mln@Rulxho.Leidenuniv.NL, or send a fax to Mrs. dr. J.G. Kruyt.

Linguistic Datasources

[From the LINGUIST List, 5-1443]

At the request of subscribers, we've been collecting the addresses of linguistic datasources which can be reached through World Wide Web. These addresses are now available to any of you who have Web access on the LINGUIST Web server at the following URL: http://engserve.tamu.edu/files/linguistics/linguist/datasource s.html

This file is also available, to those of you who read Web-LINGUIST, through the "Linguistic datasources" link. We'd be delighted to hear any comments anyone would care to make. And if there're any addresses we need to add, please let us know what they are. We'd like to emphasize that we'd be happy to include sites where individual linguists keep data they would like to make available to their colleagues. Since the Web allows us to share not merely text, but pictures and sound-recordings, we can now begin an interchange of linguistic information that is of a very different nature from that which was possible in the past.

Grammar Laboratories for the Macintosh from LINGUISTIC INSTRUMENTS, a Campus Company at the Department of Linguistics, Göteborg University.

[LINGUIST List, 5-1437]

Linguistic Instruments is a small company specializing in research instruments and teaching programs for linguists. In our series of *Grammar Laboratories* for the Macintosh we currently offer four packages:

- Phrase Structure Grammar Laboratory
- Definite Clause Grammar Laboratory
- PATR Laboratory
- Categorial Grammar Laboratory

The Grammar Laboratories are systems for writing grammars in a form that may be manipulated by a computer. They are designed as aids for students to explore formal grammars for natural language. They help the student understand the relationship between strings, rules, and trees, to grasp the concepts of parsing and generation, the notions of syntactic ambiguity and recursion, as well as other important concepts of general and computational linguistics.

For the researcher, although the grammar laboratories should not be regarded as full-fledged grammar development environments, they are nevertheless useful for testing out ideas, in a quick and simple way. Moreover, the programs are able to display analysis trees and feature structures graphically, the graphics can be formatted in all sorts of ways, and subsequently exported to other applications.

Each program has two tools, a parser and a generator. The Parser tool parses sentences and graphically displays the corresponding categories and trees (if any). The Generator tool accepts as input a start symbol and a specification of a maximal tree depth, and (randomly or systematically) generates any combination of a string, spoken utterance, category symbol, or tree.

The Grammar Laboratories form an integrated package with a generic design. Nevertheless, each laboratory has some distinguishing features:

- PSG Laboratory: A useful tool for introductory courses. It directly supports the standard notation for (context-free) phrase structure grammar, including conventions for optional and alternative constituents.

- DCG Laboratory: An environment for Definite Clause Grammar supporting variable categories, left-recursive rules, and a limited use of escape to Prolog.

- PATR Laboratory: Over and above the standard PATR formalism, this system supports list-valued features and feature structure variables. The graphical display of feature structures is enhanced with colour coding for reentrancy.

- CG Laboratory: Grammatical analyses can be displayed either in ordinary phrase structure trees or in the special kind of annotated proof trees characteristic of categorial grammar.

The Grammar Laboratories are *real* Macintosh applications, with all the functionality and user-friendliness that you have learned to expect from Macintosh programs. Each package comes with printed documentation in the form of a 20 pages booklet, as well as a collection of sample grammars.

Fully functional versions of the Grammar Laboratories, freely distributed for evaluation, can be retrieved by anonymous ftp from the following sites:

hjelmslev.ling.gu.se/pub/li/psg-laboratory-11.hqx
dcg-laboratory-11.hqx
patr-laboratory-11.hqx
cg-laboratory-11.hqx

sumex-aim.stanford.edu/info-mac/sci/psg-laboratory-11.hqx
dcg-laboratory-11.hqx
patr-laboratory-11.hqx
cg-laboratory-11.hqx

or at any mirror of info-mac.

The Grammar Laboratories are *shareware programs*. This means that if you use them, you should pay for them. For further information, please contact: Linguistic Instruments, Dept of Linguistics, Göteborg University, S-412 98 Göteborg, Sweden (Email: li@ling.gu.se)

NEW PUBLICATIONS

Les Cahiers de Grammaire

Les Cahiers de Grammaire is a journal edited by the Research Group on Syntax and Semantics (ERSS - URA 1033 CNRS) at the University of Toulouse-Le Mirail. The purpose of this publication is to act as a forum for detailed descriptions of linguistic phenomena (at syntactic, semantic, morphological or phonological level), favoring at the same time researches and discussions on the formalisms and models elaborated to account for these phenomena. *Les Cahiers de Grammaire* are also opened to pragmatic and cognitive studies grounded on descriptive and formal analyses of language. Thus, they constitute a place for exchanges and discussions between linguists, psycholinguists, philosophers of language, cogniticians and computer scientists. If you are interested in ordering some issues of *Les Cahiers de Grammaire* or in submitting some paper to be published (in French or English), you can contact Mme Laurence Lamy (address : ERSS-URA 1033 du CNRS, Maison de la Recherche, Université de Toulouse-Le Mirail, 5 allée Antonio Machado, 31058 Toulouse Cedex (Tel: (33) 61 50 46 76; Fax : (33) 61 50 36 02). The price of each issue including postage is 70FF.

Computists' Communique

Computists International is looking for members. We are a mutual-aid professional association of AI/IS/CS researchers: professors, department heads, program directors, research scientists, students, consultants, librarians, journalists, authors, programmers, and software entrepreneurs. Even an AI/HCI recruiter. We have members in more than a dozen countries and in scores of AI-related labs and departments. Ask about our new (reduced!) professional, international, and institutional rates, or sign up for the freebies below.

The Computists' Communique is our weekly email newsletter. 44 weekly issues of about 32KB (8 pages) include job ads, journal calls, NSF announcements, grant and research news, online resources, and business tips, plus current topics and tips from Internet/Bitnet/Usenet discussion groups. You get a concise, time-saving summary from dozens of sources.

A 1/4 subset is free to any who ask. (Over 900 are subscribed.) Just request the Full Moon distribution from laws@ai.sri.com. Also, please have your full name somewhere in the message. You'll get a sample newsletter, a description of the service, and then the current Communique on the day of each full moon.

Two less-condensed digests are also available. APJ covers applied and graduate-level job opportunities, plus occasional CS faculty positions outside the computational AI domain; RSW forwards research software announcements from Usenet. These "press release" digests are free to anyone, courtesy of the members of Computists International.

Remember, just ask for the Full Moon distribution and you get the Communique free once each month. No obligation or hard sell; we just want to introduce the service and keep in touch.

From: Kenneth I. Laws, Computists International (Email: laws@ai.sri.com)

The Translator
Studies in Intercultural Communication

The Translator is a refereed international journal which seeks to bring professional and academic interests closer together by addressing issues which have relevance in both academic and professional settings. Members of the editorial and advisory boards have extensive academic and professional experience. Many are accomplished translators and interpreters working at the highest levels. All have sound academic training and are involved in teaching translators and interpreters on a regular basis. Academic expertise ensures rigour and coherence in the treatment of issues; professional experience complements academic rigour with a rich store of concrete contexts and an emphasis on relevance and readability.

The Translator is not restricted in scope to any particular school of thought or academic group. Its editors and advisors hold diverse views on translation; they are nevertheless all committed to promoting a harmonious and less divisive environment in which both discipline and profession can prosper.

Order forms for The Translator will shortly be available (by mid November). To receive one, please send a note of your full POSTAL address to: The Translator, St. Jerome Publishing, 2 Maple Road West, Brooklands, Manchester, M23 9HH, U.K. Alternatively, you can fax the same details to fax number +44 0161-973-9856, or email them to mona@ccl.umist.ac.uk.

TEXT Technology
The Journal of Computer Text Processing

With volume 5 (1995), *TEXT Technology* will continue to publish articles and reviews about all facets of using computers for the creation, processing, communication, and analysis of texts. It is designed for academic and corporate researchers, writers, editors, and teachers. Each of the quarterly issues of the printed journal has about eighty 7-by-9-inch pages and perfect binding.

Submissions of articles and reviews are welcome. MLA style should be followed. Writers of book or software reviews are encouraged to contact the Editor before submitting reviews. Manuscripts (in the form of ASCII files) and inquiries should be sent by email to the Editor at JohnsonE@columbia.dsu.edu. Writers will normally receive notices about acceptance and referees' comments promptly via email. There are no page charges.

Recent issues of the quarterly journal have contained articles about text analysis, examination of the novels of Jane Austen, collation of variant texts, programming in Icon and SPITBOL-386, teaching writing using computers, as well as a directory of electronic text centers. Reviews have been published of bibliography and translation software, of Framemaker and of five new word processors including WordPerfect for DOS, for Windows, and for the Macintosh.

Plans for future issues of *TEXT Technology* include articles on text encoding (SGML and the TEI) and reviews of writing software and of Microsoft Word for Windows and for the Macintosh.

To receive subscription details and information about the price of the printed journal, send an email message to: LangnerS@columbia.dsu.edu

---

LEXIKOLOGIE - LEXICOLOGY - LEXICOLOGIE
An International Journal on the Structure of the Vocabulary

[From LINGUIST, 5-947]

We are happy to inform you that a new journal on Lexicology was founded by DE GRUYTER (Berlin - New York).

Editors: Peter Rolf Lutzeier, Guildford (Managing Editor); Dmitrij Dobrovol'skij, Moscow; Adrienne Lehrer, Tucson; Hans-Juergen Sasse, Köln

Advisory Board: Anatolij Baranov, Moscow; Volker Beeh, Düsseldorf; Leila Behrens, Cologne/Munich; Alan Cruse, Manchester; Gertrud Greciano, Strasbourg; Franz Hundsnurscher, Münster; Michael Job, Marburg; Jurij Karaulov, Moscow; and others.

Topics:

- Theory and Practice of Vocabulary Structure in Natural Languages
- The Mental Lexicon
- The Lexicon in Grammatical Theory
- The Lexicon in Computational Linguistics
- History and Methodology of Lexicology
- Reviews (in each issue)
- Info section: reports on important conferences, work-shops, etc.

"LEXICOLOGY" provides the missing forum for studies on the nature and structure of words and vocabularies. It publishes papers on theoretical issues as well as on lexical data and aims to promote research in this core area of linguistics.

As a multilingual journal LEXICOLOGY invites contributions in German, English or French. Authors are requested to send four copies of their manuscript to: Professor Dr. Peter R. Lutzeier, Department of Linguistic & International Studies, University of Surrey, Guildford, Surrey GU2 5XH, England.

2 issues per volume. First issue planned for spring/summer 1995.
Orders may be directed to Walter de Gruyter, Berlin.

For further information please contact Hans-Juergen Sasse, Universität zu Köln, Institut für Sprachwissenschaft, D-50923 Köln, Germany. (Email: am000@rs1.rrz.uni-koeln.de; Fax: +49-221-470 51 58

Sage Publications

Sage Publications is seeking to expand its line of books and journals in the areas of theoretical linguistics, applied linguistics, psycholingusitics, computational linguistics, sociolinguistics, and cognitive linguistics. If you have a book or journal idea, or have considered writing a book or starting a journal, please read on.

For a book to be of interest to Sage it must meet two criteria: 1) It must be good work based on sound scholarship, and preferably break new ground; 2) It must have a reasonable prospect of selling several thousand copies over the 3-5 year life of an edition.

The successful Sage book will draw its audience from one or more of the following sources: 1) The book might be a primary textbook for an undergraduate or graduate course; 2) The book might be supplemental reading in an upper division undergraduate or graduate course, or one of several books assigned to a seminar; 3) The book may serve as a reference work or be used for professional development, i.e., a book bought by a scholar to serve as a reference or as an introduction to a subject; 4) The book may be used as a tutorial by graduate students or for continuing education by professionals and scholars; 5) The book may be purchased by libraries (generally a function of topic relevance and a positive book review.)

Further informnation: J. Alex Schwartz, Linguistics Editor, Sage Publications, Inc., Thousand Oaks, CA 91320 (Tel: (805) 4990721; Email: Cat_McGlothlin@SAGEPUB.COM

Natural Language Software Registry

[From ELSNET listserver]

The Natural Language Software Registry (NLSR) is pleased to announce the availability of its famous Registry on the DFKI Beta Online Research 3W Server.

The URL is: http://cl-www.dfki.uni-sb.de/cl/registry/draft.html

The NATURAL LANGUAGE SOFTWARE REGISTRY is a concise summary of the capabilities and sources of language processing software available to researchers. It comprises academic, commercial, and proprietary software, with theory, specifications, and terms on which it can be acquired clearly indicated. The second edition contains about one hundred software descriptions.

If you have written a system dealing with natural language and if you want to contribute to the third version of our registry (which can make your work popular since it is distributed all over the world), you'll have to fill in the text questionnaire which is also available on the Web server. Soon, there will be an implementation of an *interactive questionnaire* in order to give you a user friendly tool to contribute to our Registry.

The document and the questionnaire are also available by anonymous ftp to

- ftp.dfki.uni-sb.de
  (directory: pub/registry),
- crlftp.nmsu.edu
  (directory: pub/non-lexical/NL_Software_Registry)
- dri.cornell.edu
  (directory: /pub/Natural_Language_Software_Registry/

or /pub/NLSR)

- svr-ftp.eng.cam.ac.uk
(directory: /pub/comp.speech/info/NaturalLanguageSoftwareRegistry)

From: Christoph Jung, Markus Vonerden, Natural Language Software Registry, Deutsches Forschungszentrum fuer Künstliche Intelligenz (DFKI), Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany (Tel: +49 (681) 303-5282; Email: registry@dfki.uni-sb.de)

Computation and Language E-Print Archive

[From CORPORA list]
You can send your paper to the "Computation and Language E-Print Archive", from where everyone will be able to retrieve it with a simple email command. Please support it by subscribing to it, submitting you papers to it, and then telling all your friends and colleagues about it. Here is how to get more information about (among other topics) submission of papers to the server, subscribing or canceling your subscription, requesting full text of any papers, retrieving macro files for these papers, searching past listings, or submitting comments to the server operators, send a message to: CMP-LG@XXX.LANL.GOV; Subject: help. Or access the hypertext help information through the WWW/Mosaic interface at "http://xxx.lanl.gov/cmp-lg/".

Selection of Messages from the ACL server CMP-LG

[From CMP-LG NEWS, November to December 1994]

TAG+3 Workshop Proceedings

An on-line proceedings for the Third International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+3), held in September 1994 in Paris, France, has been added to the cmp-lg WWW home page.

Coling '94 Workshop Proceedings

An on-line proceedings for the Fifteenth International Conference on Comptuational Linguistics (COLING '94), held in August 1994 in Kyoto, Japan, has been added to the cmp-lg WWW home page.

CMP-LG and the ACL Conferences

The Association for Computational Linguistics requests that papers submitted to ACL-95 and EACL-95 (and all ACL-sponsored conferences) be submitted as well to cmp-lg. This can be done whenever you want, as early as when the paper is submitted for reviewing or as late as after the conference, but ideally not later than when the final proceedings version is submitted. All papers will be made available not only through the normal cmp-lg mechanisms, but also through on-line proceedings. (A similar on-line proceedings is now available for ACL-94 through the WWW/Mosaic interface.)

Conference organizers wishing to have an on-line proceedings made available through cmp-lg should contact the server at cl-server@das.harvard.edu.

Anonymous FTP Retrieval and Submission

Help information on submission and retrieval of papers by anonymous FTP is now available from the server. Use the "get ftp-instructions" command or access the FTP help information from the cmp-lg hypertext help at "http://xxx.lanl.gov/cmp-lg/ftp.html".

Selection of Abstracts of Articles and Reports available, November to December 1994

Paper: cmp-lg/9411004 ((Proceedings of TAG+3, 1994))
From: egedi@unagi.cis.upenn.edu
Title: Determining Determiner Sequencing: A Syntactic Analysis for English
Author: Beth Ann Hockey, Dania Egedi, University of Pennsylvania
Comments: uuencoded compressed ps file. 4 pages.

Previous work on English determiners has primarily concentrated on their semantics or scoping properties rather than their complex ordering behavior. The little work that has been done on determiner ordering generally splits determiners into three subcategories. However, this small number of categories does not capture the finer distinctions necessary to correctly order determiners.

This paper presents a syntactic account of determiner sequencing based on eight independently identified semantic features. Complex determiners, such as genitives, partitives, and determiner modifying adverbials, are also presented. This work has been implemented as part of XTAG, a wide-coverage grammar for English based in the Feature-Based,
Lexicalized Tree Adjoining Grammar (FB-LTAG) formalism.

Paper: cmp-lg/9411005 (Proceedings of TAG+3, 1994)
From: egedi@unagi.cis.upenn.edu
Title: Constraining Lexical Selection Across Languages Using TAGs
Author: Dania Egedi, Martha Palmer, University of Pennsylvania
Comments: uuencoded compressed ps file. 4 pages.

Lexical selection in Machine Translation consists of several related components. Two that have received a lot of attention are lexical mapping from an underlying concept or lexical item, and choosing the correct subcategorization frame based on argument structure. Because most MT applications are small or relatively domain specific, a third component of lexical selection is generally overlooked - distinguishing between lexical items that are closely related conceptually. While some MT systems have proposed using a 'world knowledge' module to decide which word is more appropriate based on various pragmatic or stylistic constraints, we are interested in seeing how much we can accomplish using a combination of syntax and lexical semantics. By using separate ontologies for each language
implemented in FB-LTAGs, we are able to elegantly model the more specific and language dependent syntactic and semantic distinctions necessary to further filter the choice of the lexical item.

Paper: cmp-lg/9411006 (Proceedings of TAG+3, 1994)
From: egedi@unagi.cis.upenn.edu
Title: Status of the XTAG System
Author: Christy Doran, Dania Egedi, Beth Ann Hockey, B. Srinivas, University of Pennsylvania
Comments: uuencoded compressed ps file. 4 pages.

XTAG is an ongoing project to develop a wide-coverage grammar for English, based on the Feature-based Lexicalized Tree Adjoining Grammar (FB-LTAG) formalism. The XTAG system integrates a morphological analyzer, an N-best part-of-speech tagger, an Early-style parser and an X-window interface, along with a wide-coverage grammar for English developed using the system. This system serves as a linguist's workbench for developing FB-LTAG specifications. This paper presents a description of and recent improvements to the various components of the XTAG system. It also presents the recent performance of the wide-coverage grammar on various corpora and compares it against the performance of other wide-coverage and domain-specific grammars.

Paper: cmp-lg/9411008
From: rambow@linguist.jussieu.fr (Owen Rambow)
Title: Parsing Free Word-Order Languages in Polynomial Time
Author: Tilman Becker (U. Penn.) and Owen Rambow (Paris 7)
Comments: 4 pages, uuencoded compressed ps file
Report-no: In 3e Colloque International sur les Grammaires
d'Arbres Adjoints (TAG+3). Technical Report TALANA-RT-94-01, TALANA, Université Paris 7, 1994.

We present a parsing algorithm with polynomial time complexity for a large subset of V-TAG languages. V-TAG, a variant of multi-component TAG, can handle free-word order phenomena which are beyond the class LCFRS (which includes regular TAG). Our algorithm is based on a CYK-style parser for TAGs.

Paper: cmp-lg/9411010
From: Jean.Gaschler@xerox.fr (Jean Gaschler)
Title: The "Whiteboard" Architecture: a way to integrate heterogeneous components of NLP systems

Author: Christian Boitet (GETA, IMAG (UJF & CNRS) & Mark Seligman (ATR Interpreting Telecommunications Research Labs)
Comments: Postscript, 6 pages, COLING-94\\

We present a new software architecture for NLP systems made of heterogeneous components, and demonstrate an architectural prototype we have built at ATR in the context of Speech Translation.

Paper: cmp-lg/9411018
From: atle@babel.ifl.uib.no (Atle Ro)
Title: Interlanguage Signs and Lexical Transfer Errors
Author: Atle Ro, Department of Phonetics and Linguistics, University of Bergen
Comments: Paper presented at COLING-94. 4 pages, compressed Postcript file.

A theory of interlanguage (IL) lexicons is outlined, with emphasis on IL lexical entries, based on the HPSG notion of lexical sign. This theory accounts for idiosyncratic or lexical transfer of syntactic subcategorisation and idioms from the first language to the IL. It also accounts for developmental stages in IL lexical grammar, and grammatical variation in the use of the same lexical item. The theory offers a tool for robust parsing of lexical transfer errors and diagnosis of such errors.

Paper: cmp-lg/9411022
From: "David D. Palmer" <dpalmer@CS.Berkeley.EDU>
Title: Adaptive Sentence Boundary Disambiguation
Author: David D. Palmer (University of California, Berkeley) and Marti A. Hearst (Xerox PARC)
Comments: 6 pages, uuencoded compressed Postscript
Report-no: ACL Proceedings, 4th Applied Natural Language Processing, 13-15 October 1994, Stuttgart

Labeling of sentence boundaries is a necessary prerequisite for many natural language processing tasks, including part-of-speech tagging and sentence alignment. End-of-sentence punctuation marks are ambiguous; to disambiguate them most systems use brittle, special-purpose regular expression grammars and exception rules. As an alternative, we have developed an efficient, trainable algorithm that uses a lexicon with part-of-speech probabilities and a feed-forward neural network. After training for less than one minute, the method correctly labels over 98.5\% of sentence boundaries in a corpus of over 27,000 sentence-boundary marks. We show the method to be efficient and easily adaptable to different text genres, including single-case texts. (Code from latest version of the program described is available from dpalmer@cs.berkeley.edu)

Paper: cmp-lg/9411023
From: ono@eel.rdc.toshiba.co.jp
Title: Abstract Generation based on Rhetorical Structure Extraction
Author: Kenji Ono, Kazuo Sumita, Seiji Miike Research and Development Center, Toshiba Corporation Komukai-Toshiba-cho 1, Saiwai-ku, Kawasaki, 210, Japan
Comments: 5 pages including 2 eps Figure, using epsbox.sty, art10.sty Report-no: COLING-94, pp.344 - 348.

We have developed an automatic abstract generation system
for Japanese expository writings based on rhetorical structure extraction. The system first extracts the rhetorical structure,
the compound of the rhetorical relations between sentences, and then cuts out less important parts in the extracted structure
to generate an abstract of the desired length. Evaluation of the generated abstract showed that it contains at maximum 74\% of the most important sentences of the original text. The system is now utilized as a text browser for a prototypical interactive document retrieval system.

Paper: cmp-lg/9411025
From: Gregor Erbach <erbach@CoLi.Uni-SB.DE>
Title: Multi-Dimensional Inheritance
Author: Gregor Erbach (University of the Saarland, Computational Linguistics Dept.)

Comments: 9 pages, styles: a4,figfont,eepic,epsf
Report-no: CLAUS Report 40

In this paper, we present an alternative approach to multiple inheritance for typed feature structures. In our approach, a feature structure can be associated with several types coming from different hierarchies (dimensions). In case of multiple inheritance, a type has supertypes from different hierarchies. We contrast this approach with approaches based on a single type hierarchy where a feature structure has only one unique most general type, and multiple inheritance involves computation of greatest lower bounds in the hierarchy. The proposed approach supports current linguistic analyses in constraint-based formalisms like HPSG, inheritance in the lexicon, and knowledge representation for NLP systems. Finally we show that multi-dimensional inheritance hierarchies can be compiled into a Prolog term representation, which allows to compute the conjunction of two types efficiently by Prolog term unification.

Paper: cmp-lg/9411026
From: Jean.Gaschler@xerox.fr (Jean Gaschler)
Title: Manipulating Human-oriented Dictionaries with very simple tools
Author: Jean Gaschler & Mathieu Lafourcade (GETA, IMAG (UJF & CNRS)
Comments: uuencoded gzipped Postscript, 4 pages, COLING-94

This paper presents a methodology for building and manipulating human-oriented dictionaries. This methodology has been applied in the construction of a French-English-Malay dictionary which has been obtained by "crossing" semi-automatically two bilingual dictionaries. We use only Microsoft Word, a specialized language for writing transcriptors and a small but powerful dictionary tool.

Paper: cmp-lg/9411014
Title: Automatically Identifying Morphological Relations in =
Machine-Readable Dictionaries
Author: Joseph Pentheroudakis and Lucy Vanderwende, Microsoft Corporation
Comments: PostScript, 19 pages, 250kb; from Proceedings of the 9th Annual Conference of the UW Centre for the New OED and Text Research, 1993
Report-no: MSR-TR-93-06

Paper: cmp-lg/9411028
From: dmc@cam.sri.com (David Carter)
Title: The Speech-Language Interface in the Spoken Language Translator
Author: David Carter and Manny Rayner (SRI International, Cambridge)
Comments: 9 pages, LaTeX. Published: Proceedings of TWLT-8, December 1994.
Report-no: CRC-051

The Spoken Language Translator is a prototype for practically useful systems capable of translating continuous spoken language within restricted domains. The prototype system translates air travel (ATIS) queries from spoken English to spoken Swedish and to French. It is constructed, with as few
modifications as possible, from existing pieces of speech and language processing software. The speech recognizer and language understander are connected by a fairly conventional pipelined N-best interface. This paper focuses on the ways in which the language processor makes intelligent use of the sentence hypotheses delivered by the recognizer. These ways include (1)
producing modified hypotheses to reflect the possible presence of repairs in the uttered word sequence; (2) fast parsing with a version of the grammar automatically specialized to the more frequent constructions in the training corpus; and (3) allowing syntactic and semantic factors to interact with acoustic ones in the choice of a meaning structure for translation, so that the
acoustically preferred hypothesis is not always selected even if it is within linguistic coverage.

Paper: cmp-lg/9411022
Title: Adaptive Sentence Boundary Disambiguation

Author: David D. Palmer and Marti A. Hearst
Comments: This is a Latex version of the previously submitted ps file (formatted as a uuencoded gz-compressed .tar file created by csh script). The software from the work described in this paper is available by contacting dpalmer@cs.berkeley.edu

Paper: cmp-lg/9411027
Title: Classifier Assignment by Corpus-based Approach
Author: Virach Sornlertlamvanich, Wantanee Pantachat and Surapant
Meknavin, Linguistics and Knowledge Science Laboratory (LINKS), National Electronics and Computer Technology Center (NECTEC), Bangkok, Thailand
Comments: 6 pages, compressed Postscript file
Report-no: COLING-94, Vol.1, pp.556-561.

Paper: cmp-lg/9411030
From: srini@unagi.cis.upenn.edu
Title: Complexity of Scrambling: A New Twist to the Competence - Performance Distinction
Author: Aravind K Joshi, University of Pennsylvania
Comments: uuencoded compressed ps file. 4 pages.

In this paper we discuss the following issue: How do we decide whether a certain property of language is a competence property or a performance property? Our claim is that the answer to this question is not given a-priori. The answer depends on the formal devices (formal grammars and machines) available to us for describing language. We discuss this issue in the context of the complexity of processing of center embedding (of relative clauses in English) and scrambling (in German, for example) from arbitrary depths of embedding.

Paper: cmp-lg/9411031
From: Ehud Reiter <ehud@cogentex.com>
Title: Automatic Generation of Technical Documentation
Author: Ehud Reiter (CoGenTex, Ithaca, USA), Chris Mellish (University of Edinburgh, UK), and John Levine (University of Edinburgh, UK)
Comments: uuencoded compressed tar file, with LaTeX source and ps figures. Will appear in APPLIED ARTIFICIAL INTELLIGENCE journal, volume 9 (1995).

Natural-language generation (NLG) techniques can be used to automatically produce technical documentation from a domain knowledge base and linguistic and contextual models. We discuss this application of NLG technology from both a technical and a usefulness (costs and benefits) perspective. This discussion is based largely on our experiences with the IDAS documentation-generation project, and the reactions various interested people from industry have had to IDAS. We hope that this summary of our experiences with IDAS and the lessons we have learned from it will be beneficial for other researchers who wish to build technical-documentation generation systems.

Paper: cmp-lg/9412003
From: Joerg Ueberla <ueberla@signal.dra.hmg.gb>
Title: An Extended Clustering Algorithm for Statistical Language Models
Author: Joerg P. Ueberla (Forum Technology - DRA Malvern)
Comments: 27 pages, latex, comments welcome
Report-no: DRA/CIS(CSE1)/RN94/13

Statistical language models frequently suffer from a lack of training data. This problem can be alleviated by clustering, because it reduces the number of free parameters that need to be trained. However, clustered models have the following drawback: if there is ``enough'' data to train an unclustered model, then the clustered variant may perform worse. On currently used language modeling corpora, e.g. the Wall Street Journal corpus, how do the performances of a clustered and an unclustered model compare? While trying to address this question, we develop the following two

ideas. First, to get a clustering algorithm with potentially high performance, an existing algorithm is extended
to deal with higher order N-grams. Second, to make it possible to cluster large amounts of training data more efficiently, a heuristic to speed up the algorithm is presented. The resulting clustering algorithm can be used to cluster trigrams on the Wall Street Journal corpus and the language models it produces can compete with existing back-off models. Especially when there is
only little training data available, the clustered models clearly outperform the back-off models.

Paper: cmp-lg/9412004
From: Marc Light <light@cs.rochester.edu>
Title: Knowledge Representation for Lexical Semantics: Is Standard First Order Logic Enough?
Author: Marc Light and Lenhart Schubert, University of Rochester
Comments: Presented at the "Future of the Dictionary" workshop, Grenoble, France (October, 1994), 12 pages (uuencoded compressed postscript).

Natural language understanding applications such as interactive planning and face-to-face translation require extensive inferencing. Many of these inferences are based on the meaning of particular open class words. Providing a representation that can support such lexically-based inferences is a primary concern of lexical semantics. The representation language of first order logic has well-understood semantics and a multitude of inferencing systems have been implemented for it. Thus it is a prime candidate to serve as a lexical semantics representation. However, we argue that FOL, although a good starting point, needs to be extended before it can efficiently and concisely support all the lexically-based inferences needed.

Paper: cmp-lg/9412006
From: Ted Briscoe <ejb@linc.cis.upenn.edu>
Title: Robust stochastic parsing using the inside-outside algorithm
Author: Briscoe, Ted and Waegner, Nick (University of Cambridge)
Comments: Revised and updated version of paper from AAAI Workshop on Probabilistically-based Natural Language Processing Techniques, 1992, 16 pages, uuencoded, compressed postscript.

The paper describes a parser of sequences of (English) part-of-speech labels which utilises a probabilistic grammar trained using the inside-outside algorithm. The initial (meta)grammar is defined by a linguist and further rules compatible with metagrammatical constraints are automatically generated. During training, rules with very low probability are rejected yielding a wide-coverage parser capable of ranking alternative analyses. A series of corpus-based experiments describe the parser's performance.

PUBLICATIONS NOTIFIED

Centre for Cognitive Science and Human Communication Research Centre, University of Edinburgh

The Centre for Cognitive Science (CCS), established in 1969 as the 'School of Epistemics', has had from its beginning the aim of contributing to the development of cognitive science through both research and postgraduate training. The research programme of the Centre is directed mainly towards the syudy of language and related cognitive processes, mainly from a computational perspective. The range of research projects in CCS and the Human Communication Research Centre (HCRC) is broadly representative of current interests in cognitive science.

The HCRC is ESRC's centre of excellence in the interdisciplinary study of cognition and computation in human communication. Drawing together researchers from Edinburgh, Glasgow and Durham, HCRC focuses on the psychological aspects of real langauge

processing. The HCRC is closely associated with CCS; the two contribute towards a joint research environment.

Much of the two centres' research is carried out within funded research projects, with funding from the Engineering and Physical Sciences Research Council, the Economic and Social Research Council, the Medical Research Council, and the European Community ESPRIT and LRE Programmes. Current funded projects include: Integrated Language Database. – EuroTac: Methodologies for Constructing Knowledge Bases for Natural Language Processing Systems. – DANDELION: Discourse Functions and Discourse Representation. – Towards a Declarative Theory of Discourse. – DYANA: Dynamic Interpretation of Natural Language. – MENELAS: an Access System for Medical Records using Natural Language. – The Reusability of Grammatical Resources. –SISTA: Semi-automatic Indexing System for Technical Abstracts. – The Concept of Discourse Style. – A Psychologically Relevant Model of Belief. – Novel Tools for Modelling Memory Processes. – Real Time Language Generation. – SIGNAL: Specificity of Information in Graphics and Natural Language. –Automatic Evaluation of the Quality of Computer-generated Text. – Cognitive Evaluation of Hyperproof: a Graphical Program for Teaching Logic. – WOPIS Word order, Prosody and Information Structure. – Designing systems of coupled neural networks. – The construction of networks of neurons and the computations performed by them. –Connectionist modelling of spoken word recognition. –QUADS: Quantification and Dynamic Semantics. – FraCaS: Framework for Computational Semantics.

Recent research papers, 1993-1994:

Centre for Cognitive Science

RP55: Andreas Schoter, Compiling feature structures into terms: an empirical study in Prolog (1993) £1.80
RP56: Steven Bird and Ewan Klein, Enriching HPSG phonology (1993) £1.50
RP57: Steven Bird and Oliver Stegen, Tone in the Bamileke Dschang associative construction (1993) £1.00
RP58: Toby Tyrell, Defining the action selection problem (1994) £0.50
RP59: Mark J.L.Orr, Regularised centre recruitment in radial basis fuction networks (1994) £0.70
RP60: Martin W.Simmen, Multidimensional scaling of binary dissimilarities: direct and derived approaches (1994) £0.80
RP61: Terry Myers, History of the Centre for Cognitive Science (1994) £0.50
RP62: Adam Przepiórkowski, Critical review of approaches to multiple wh-movement (1994) £1.00
RP63: Geoffrey J.Goodhill et al., An evaluation of the use of multidimensional scaling for understanding brain connectivity (1994) £1.00
RP64: Andreas Schröter, Evidential bilattice logic and lexical inference (1994) £1.10
RP65: Massimo Poesio, Weak definites (1994) £1.20

Further information: The Librarian, Centre for Cognitive Science, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, UK. (Fax: +44-131-650-4587; Email: librarian@cogsci.ed.ac.uk)

Human Communication Research Centre

Recent research papers, 1993-1994:

RP-40: Jim Miller and Regina Weinert, The function of LIKE in dialogue (January 1993) £1.10
RP-41: Regina Weinert and Jim Miller, Cleft constructions in spoken dialogue (February 1993) £1.30
RP-42: Robin Cooper, A note on the relationship between linguistic theory and linguistic engineering (February 1993) £0.70
RP-43: Jean Carletta, Richard Caley and Stepehn Isard, A system architecture for simulating time-constrained language production (February 1993) £0.70
RP-44: Enric Vallduvi, Information packaging: a survey (August 1993) £1.60
RP-45: J.A.Taylor, An ATMS-based belief model for dialogue simulation (October 1993) £1.10

RP-46: Jonathan Ginzburg, Resolving questions (October 1993) £1.90
RP-49: Enric Vallduvi and Ron Zacharski, Accenting phenomena, association with focus, and the recursiveness of focus-ground (June 1994) £1.10
RP-50: Andrew John Merrison, Anne H.Anderson and Gwyneth Doherty-Sneddon, An investigation into the communicative abilities of aphasic subjects in task oriented dialogue (June 1994) £1.10
RP-51: Keith Stenning, Logic as a foundation for a cognitive theory of modality assignment (June 1994) £1.00
RP-52: E.G.Bard, D.Robertson, and A.Sorace, Magnitude estimation of linguistic acceptability (July 1994) £1.60
RP-54: Rosemary J.Stevenson, Rosalind A.Crawley and David Kleinman, Thematic roles, focus and the representation of events (August 1994) £1.20
RP-55: Enric Vallduvi, Polarity items, n-words, and minimizers in Catalan and Spanish (August 1994) £1.20
RP-56: Richard Shillcock and Paul Cairns, Neglect dyslexia and the neuropsychology of reading: a connectionist model of neglect dyslexia with no attentional component (July 1994) £0.70
RP-58: Rosemary J.Stevenson and Agnieszka Urbanowicz, The effects of sentence subject, initial mention and pragmatic plausibility on the accessibility of a pronoun's antecedent (November 1994) £1.20
RP-59: Corin Gurr, Supporting formal reasoning for safety critical systems (August 1994) £0.70
RP-60: Jasper Taylor, Using hierarchical autoepistemic logic to model beliefs in dialogue (November 1994) £1.30
RP-62: Francis Corblin, Multiple negation processing (November 1994) £1.30

Further information: Fiona-Anne Malcolm, Human Communication Research Centre, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, UK. (Fax: +44-131-650-4445; Email: fianne@cogsci.ed.ac.uk)

Working Papers from Centre for Cognitive Science:
WP vol.1: Nicholas Haddock, Ewan Klein and Glyn Morrill (eds) Categorial grammar, unification grammar and parsing (February 1987) £7.75
WP vol.2: Jon Oberlander (ed) Temporal reference and quantification: an IQ perspective (September 1987) £5.00
WP vol.3: Nick Braisby and Richard Cooper (eds) Situation theoretic studies in psychology, language and logic (July 1989) £4.25
WP vol.4: George Dunbar, Bradley Franks and Terry Myers (eds) Papers from the 1989 Edinburgh Round Table on the Mental Lexicon (June 1989) £5.25
WP vol.5: Guy Barry and Glyn Morrill (eds) Studies in categorial grammar (May 1990) £5.25
WP vol.6: Elisabet Engdahl, Mike Reape, Martin Mellor and Richard Cooper (eds) Parametric variation in Germanic and Romance (July 1990) £8.00
WP vol.7: Steven Bird (ed) Declarative perspectives in phonology (October 1991) £7.75
WP vol.8: T.Mark Ellison and James M.Scobbie (eds) Computational phonology (October 1993) £7.50
WP vol.9: Catrin Sian Rhys, David Adger and Ana von Klopps (eds) Functional categories, arguments structure and parametric variation (November 1994) £7.75

Verbmobil Reports

The following reports of the Vermobil project are available from Reinhart Karger, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Stuhlsatzenhausweg 3, Bau 43.1, D-66123 Saarbrücken, Germany (Email: karger@dfki.uni-sb.de)

1. Wolfgang Wahlster: Verbmobil: Übersetzung von Verhandlungsdialogen (July 1993).
2. Dafydd Gibbon: Generalised DATR for flexible lexical access: Prolog specification (October 1993).

3. Christel Kemke: Evaluation von Wissensrepräsentationsformalismen (November 1993).
4. R. Kompe, A. Batliner, A. Kießling, U. Kilian, H. Niemann, E. Nöth, P. Regel-Brietzmann: Automatic classification of prosodically marked phrase boundaries in German (1993).
5. Susanne Jekat-Rommel: Aspekte des Dolmetschverhaltens (1993).
6. Johan Bos, Elsbeth Mastenbroek, Scott McGlashan, Sebastian Millies, Manfred Pinkal: The Verbmobil semantic formalism (Version 1.3).
7. Miriam Butt, Sabine Reinhard, Heike Winhart: Korpusanalyse und Auswahl der Teilphänomene (February 1994).
8. Jan Amtrup: Transfer and architecture: views from chart parsing (March 1994).
9. Andreas Hauenstein, Hans Weber: An investigation of tightly coupled time synchronous speech language interfaces using a unification grammar (March 1994).
10. Markus Duda: Lexicon access on parallel machines (April 1994).
11. Folker Caroli, Rita Nübel, Bärbel Ripplinger, Jörg Schütz: Transfer in Verbmobil (May 1994).
12. Herbert Tropf: Spontansprachliche syntaktische Phänomene: Analyse eines Korpus aus der Domäne "Terminabsprache" (April 1994).
13. Günther Görz, Marcus Kesseler: Anytime algorithms for speech parsing? (February 1994).
14. Hagen Langer: DATR without nodes and global inheritance (April 1994).
15. Kai Hübener, Julie Carson-Berndsen: Phoneme recognition using acoustic events (June 1994).
16. Stefan Wermter, Volker Weber: Learning fault-tolerant speech parsing with SCREEN (June 1994).
17. Markus Duda: From DATR to PATR via DUTR - an interface formalism (July 1994).
18. Kai Lebeth: Morphosyntaktischer Strukturaufbau (July 1994).
19. Stefan Geissler: Erläuterungen zur Umsetzung einer HPSG im Basisformalismus STUF III (July 1994).
20. Stefan Geissler: Lexikalische Regeln in der IBM-Basisgrammatik (July 1994).
21. Marc Hüsken, Tibor Kiss: Some remarks on the grammar of date and time (July 1994).
22. K. Harbusch, G. Kikui, A. Kilger: Default handling in incremental generation (July 1994).
23. Shinichi Kameyama, Ilona Maleck: Konstellation und Szenario von Terminabsprachen (August 1994).
24. Marcus Kesseler: Distributed control in Verbmobil (August 1994).
25. Johan Bos: Presupposition as anaphora in the Verbmobil semantic formalism (August 1994).
26. Wolfgang Menzel: Parsing of spoken language under time constraints (August 1994).
27. Shinichi Kameyama, Ilona Maleck: Transkriptionen authentischer und simulierter Termin absprachen (July 1994).
28. Birte Schmitz, Susanne Jekat-Rommel: Eine zyklische Approximation an Sprechhandlungstypen -- zur Annotierung von Äußerungen in Dialogen (September 1994).
29. Joachim Quantz, Manfred Gehrke, Uwe Küssner, Birte Schmitz: The VERBMOBIL domain model version 1.0 (September 1994).
30. Frank Keller: Extraposition in HPSG (September 1994).
31. A.Kießling, R.Kompe, H.Niemann, E.Nöth, and A.Batliner: "Roger", "Sorry", "I'm still listening": dialog guiding signals in information retrieval dialogs (October 1994).
32. J.Denzler, R.Kompe, A.Kießling, H.Niemann, and E.Nöth: Going back to the source: inverse filtering of the speech signal with ANNs (October 1994).
33. H.Niemann, J.Denzler, B.Kahles, R.Kompe, A.Kießling, E.Nöth, V.Strom: Pitch determination considering laryngealization effects in spoken dialogs (October 1994).
34. H.Niemann, E.Nöth, E.G. Schukat-Talamazzini, A.Kießling, R.Kompe, T.Kuhn, S.Rieck: Phonetic and prosodic analysis of speech (October 1994).
35. A.Kießling, R.Kompe, A.Batliner, H.Niemann, and E.Nöth: Automatic labeling of phrase accents in German (October 1994).
36. A.Kießling, R.Kompe, H.Niemann, E.Nöth, A.Batliner: Detection of phrase boundaries and accents (October 1994).
37. G.Bakenecker, U.Block, A.Batliner, R.Kompe, E.Nöth, P.Regel-Brietzmann: Improving parsing by incorporating prosodic 'sentence breaks' into a grammar (September 1994).

38. Julie Carson-Berndsen, Martina Pampel: Diagnostic evaluation in linguistic word recognition (August 1994).

PUBLICATIONS RECEIVED

Journals

AAMT Journal *no.7 (June 1994).* p.1: Machine translation as a tool (Masahiro Nagasawa). -- pp.3-5: BehaviorTran: its current status, prospects and philosophy (Jing-Shin Chang and Keh-Yih Su). --pp.6-12 : User-oriented machine translation [Toptran. See ? in this issue]. -- pp.13-16 : ATLAS machine translation service offered on PCCs. -- pp.17-19, 26: Machine translation using a statistical approach (Masaaki Nagata). -- pp.20-22: PC-VAN IBS machine translation service by NEC Corporation. -- pp.23: Bravice J/E ver 5.1 by MT Laboratories Co.Ltd. -- pp.24: PENSEE for Windows. -- pp.25-26: ASTRANSAC by Toshiba Corporation. -- pp.27-29: Current MT activities in Europe (Ralf Steinberger).
no.8 (September 1994). p.1: Visions for future MT (Takashi Nonouchi). -- pp.3-5: Coling94 tutorial sessions "Future directions of MT - language, meanings and translation". -- pp.6-10: Dictionary for machine translation (Toru Yoda). -- p.11: Central Research Laboratory, Hitachi Ltd. (Hiroyuki Kaji). --pp.12-13: HICATS by Hitachi Co., Ltd. -- pp.14-15: TransLand Mac edition: Japanese-English translation software by Brother Industries Ltd. -- pp.16-17: ATLAS/Win English-Japanese and Japanese-English machine translation system by Fujitsu Ltd. --pp.18-19: PC-Transer/ej for Windows [from Nova]. -- pp.20-21: Honyaku kobo: Panasonic English to Japanese automatic translating system. -- pp.22-24: Digests [of survey reports].

Computational Linguistics, *vol.20, no.3 (? 1994). Special issue on Computational Phonology.* pp.331-378: Regular models of phonological rule systems (Ronald M.Kaplan and Martin Kay). -- pp.381-417: The reconstruction engine: a computer implementation of the comparative method (John B.Lowe and Martine Mazaudon). --pp.421-451: The acquisition of stress: a data-oriented approach (Walter Daelemans, Steven Gillis, and Gert Durieux). -- pp.455-491: Phonological analysis in typed feature systems (Steven Bird and Ewan Klein).

Elsnews, *vol.3 no.3 (September 1994)* Contents include: pp.2-3: Corpus-based methods - second European summer school [Utrecht, July 1994]. -- pp.4-5: Speech maps: exploring new territory (Christian Abry and Pierre Badin). -- p.7: The European multilingual corpus: a user's view (Kenneth Ward Church). --pp.8-9: Language Engineering Convention '94 (Paris, 6-8 July) (Nicolas Nicolov).

INL Infoterm Newsletter *71/72 (March-June 1993).*

Language Industry Monitor *no.22 (July-August 1994)* Contents include: pp.2-6: Trados: ten years on. -- pp.7-9: The Sietec connection. -- pp.9-10: XLT ready for liftoff [Xerox Lexical Technology]. -- pp.11-12: Globalink and MicroTac tie the knot.

Language International, *vol.6 no.5 (October 1994).* Contents include: pp.4-5: Language engineers meet in Paris (Geoffrey Kingscott). -- pp.12: IBM personal dictation system (Sabine Nixon). -- pp.14-15: Systran on PC imminent (Geoffrey Kingscott). *vol.6 no.6 (December 1994).* Contents include: pp.4-7: TranslationManager for Windows (John Newton). -- pp.8-10: Universal Word for Windows (Quentin Reid).

LISA Forum Newsletter, *vol.3 no.3 (August 1994).* Contents include: pp.1-5: Confusion in the cost-value chain of localization services (Jaap van der Meer). -- pp.5-8: Quality issues in internationalization (I18N) and localization (L10N) (Emmanuel Uren, Robert Howard and Tiziana Perinotti). -- pp.10-12: Translation tool technology - a WHA evaluation report [TranslationManager/2 and Text Translation Tool] (Emilio Alesiani).

Machine Translation, *vol.9 no.1 (1994).* Contents: pp.1-19: The lexical unit in the Metal MT system (Thierry Fontenelle, Geert Adriaens, and Gert de Braekeleer). -- pp.21-59: A model of comparative stylistics for machine translation (Chrysanne DiMarco and Keith Mah). -- pp.61-80: Machine assisted translation from English to a Slavic language: what linguistics and programming methodologies can do for it.

Terminology Standardization and Harmonization: newsletter of ISO/TC 37, *vol.6 no.2 (June 1994)*

Tribune des Industrie de la Langue et de l'Information Electronique, *no.15 (1994)*. [Issue devoted to] Gestion electronique de documents.

Books

Blekhman, M.S. et al.: **Description of Russian and Ukrainian morphologies in commercial machine translation systems**. Kharkov, 1994. 41pp.

Ramm, Wiebke (ed.): **Text and context in machine translation: aspects of discourse representation and discourse processing**. (Studies in Machine Translation and Natural Language Processing, vol.6) Luxembourg: European Comission, 1994. 150pp. ISSN 1017-6568.

Conference proceedings

Proceedings of the Evaluators' Forum, April 21st-24th, 1991, Les Rasses, Vaud, Switzerland. Edited by Kirsten Falkedal. Carouge (Geneva): ISSCO, 1994. x,253 pp. [For a report of the conference see MTNI#1: 10]

MT Evaluation: basis for future directions. Proceedings of a workshop sponsored by the National Science Foundation, 2-3 November 1992, San Diego, California. Washington, DC.: Association for Machine Translation in the Americas [and] International Association for Machine Translation, 1994. var.pp. [For a report of the conference see MTNI#4: 16-18]

Proceedings of the North Texas Natural Language Processing Workshop, NTNLP-94, May 23, 1994, University of Texas at Arlington. iii,104 pp.

International Conference on New Methods in Language Processing (NeMLaP). Proceedings of the conference, September 14-16 1994, The University of Manchester Institute of Science and Technology. Editor: Daniel Jones. [Manchester]: Centre for Computational Linguistics [UMIST], 1994. vii,261 pp.

Technology partnerships for crossing the language barrier. Proceedings of the First Confernce of the Association for Machine Translation in the Americas, 5-8 October 1994, Columbia, Maryland, USA. 246pp. [For a report of the conference see ? in this issue.]

Fourth Conference on Applied Natural Language Processing. Proceedings of the conference, 13-15 October 1994, Stuttgart, Germany. San Francisco: Morgan Kaufmann for ACL, 1994. x,216pp. ISBN: 1-800745-7323.

Proceedings of the Workshop on Compound Nouns: Multilingual Aspects of Nominal Composition, 2-3 December 1994, Geneva, Switzerland. Edited by Pierrette Bouillon and Dominique Estival. Genève: ISSCO. 189pp.

Reports

Final evaluation of the results of Eurotra: a specific programme concerning the preparation of the development of an operational Eurotra system for Machine Translation. (COM (94) 69 final). Brussels: Commission of the European Communities, Sept 1994.

Linguistic Research & Engineering (LRE): an overview, June 1994. Editors: Roberto Cencioni and Ewald Klein. University of Edinburgh, 1994.

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.