Issue No.18
October/November 1997
ISSN 0965-5476
Published in February, June, and October

MT News International

Newsmagazine of the International Association for Machine Translation

In this issue...

[Note: This electronic version contains items omitted from the printed version.]

From the editor

With much regret I have to announce that from this issue of MTNI I must relinquish my editorship -- it is not possible to be simultaneously editor of the newsletter and an IAMT officer. I should like therefore to thank all those who have supported me over the past six years, particularly my fellow editors in the regional associations. Special words of gratitude must go to those who have put in so much hard word on the printing and publishing side: Joseph Pentheroudakis (issues 1 to 10) and Jane Zorrilla (issues 11 to the present). Thanks also are due to all contributors, whose prompt submissions have enabled me to send off every issue for printing at the end of the month before publication. Others who have contributed greatly behind the scenes, but have not been formally acknowledged in the Editorial panels, have been Muriel Vasconcellos and Colin Brace, particularly during the last two years. Although I shall no longer be editor, it is my hope that I can continue to write contributions of various kinds to future issues of the newsletter. I wish my successor every success in the continuation of a publication which has, I believe, not only established itself as the principal forum for the International Association for Machine Translation but which is now also seen as a major source of information about contemporary developments in machine translation and computer-aided translation systems throughout the world.

John Hutchins
September 1997

SPOTLIGHT ON THE NEWS

Microsoft to Acquire a 20 Percent Share in TRADOS

[Press release]
REDMOND, Wash., Sept. 9 -- Microsoft Corporation today announced it plans to acquire a 20 percent minority share of TRADOS GmbH, of Stuttgart, Germany, in a move designed to accelerate the delivery of localized software products.

TRADOS products are utilized extensively by the software industry to create foreign-language "localized" software products. Microsoft ships its products in more than 30 languages, and the investment in TRADOS reflects a commitment to quickly and efficiently deliver localized products internationally.

"Investing in TRADOS technology will secure our ability to produce high-quality localized products in the most cost-effective manner," said Franz Rau, director of internal tools at Microsoft. "This investment will allow us to work with TRADOS to keep in step with Microsoft's evolving needs."

Microsoft plans to use TRADOS software as its internal localization memory store. This will allow Microsoft to more effectively reuse already localized text from product to product. The minority share in TRADOS further solidifies the relationship between the two companies. As part of the agreement, Microsoft has also signed a long-term support-and-development contract to service Microsoft-specific needs. TRADOS expects this move will enable it to step up the pace for new product development.

"Our relationship with Microsoft will allow TRADOS to consolidate and expand its position as a worldwide leader in translation tools," said Iko Knyphausen, CEO of TRADOS.

"Having the benefit of a strong investor is important to us at this stage of the development of the company," said Jochen Hummel, president of TRADOS. "We'll be able to take advantage of Microsoft's experience and keep up our 100 percent growth rate of the past three years."

For more information, visit http://www.trados.com/, or call TRADOS Corp. in the United States at 703-683-6900. In Europe, call TRADOS GmbH at 49-711-168-770.

PRODUCTS AND SYSTEMS

Linguistics Products Improves PC-TRANSLATOR

PC-TRANSLATOR is now available under Microsoft Windows, and (when supplied with sixteen of the available eighteen language pairs) on a single CD. Its dictionary editor is dramatically improved and its speed, with a Pentium PC, is much faster. It is also more user-friendly, and a substantial price reduction has made it far more affordable, while the need to consult its manual has almost been eliminated with Windows Help Screens. One of PC-TRANSLATOR's features, its ability to preserve the formatting of WordPerfect and Microsoft Word (RTF) text files, can produce similar results when it is used to translate Windows Help Screens.

Developments of PARS systems

Michael Blekhman

From English into East Slavonic languages

Translating Internet files

Lingvistica '93 Co. have come up with the new versions of the bidirectional PARS and PARS/U MT systems, for translating between English and Russian and English and Ukrainian, respectively. The first thing to point out is that PARSes now translate not only WinWord, but also HTML files. Correspondingly, the user selects either the "MS Word" or the "HTML" mode in PARS options. To have an HTML text translated, he/she displays the text using, for example, Netscape Navigator, copies the text portion to be translated to the clipboard, and activates PARS. The system main window is displayed, in which the user may set up the combination of dictionaries to be used in the translation session. It is important that the system saves the configuration, so there is no need to make the settings in each translation session: PARS will behave the way it did before.

The translation lasts for about 2-3 sec. per page (250 words) on a Pentium-type computer, and the target text is pasted in a special under the source text. The polysemantic words are marked with asterisks in the target text, which is one of PARS "visiting cards": a double click on the asterisk displays translation options, another double click on a more appropriate variant substitutes the initial one in the text. At the same time, if the user does not want to see the asterisks in the translation, he/she may simply disable the "translation variants" mode. The target text may be saved in a file for further editing, if required. Another point of importance is that PARSes can translate screen Helps in the same way.

Both PARSes are marketed in North America: PARS by Virginia-based POLYGLOSSUM, Inc., and PARS/U by Erudite Corp., Toronto, Canada.

As to the ex-Union market, PARS is becoming one of the most (if not the most) popular systems in Russia. It is distributed by Moscow-based ETS Publishers Ltd. on CD-ROM together with Polyglossum dictionaries. The disks are extremely cheap, which makes them accessible to broad public. The prices are between $15 to $30, and the disks are sold practically at each Moscow computer and software shop. In 1996, about 10,000 disks were sold and placed at dealers.

Lexicographic work

PARS is a very large project aimed at covering all subject areas in science and technology. Presently, PARS terminological dictionaries comprise over 700,000 word-entries in each part, English-Russian and Russian-English. Dictionaries on mathematics (80,000), automobile building (50,000), and chemistry have been added to PARS using Lingvistica '93 know-how, which includes automatic assigning grammatical information to Slavonic words.

Also, the PARS/Avia project is under way, on the order of the Ukrainian giant, O.Antonov Aviation Design Bureau. The task is to make PARS capable of producing draft translations of aviation documentation from Russian into English and save human efforts at least 2-3 times. The first stage of the project (to be finished in April) consists in compiling a set of specialized dictionaries on various aspects of aircraft engineering.

The second stage will include developing a large documentation management system that will comprise PARS as a translation module. Other modules will be a database management system, an OCR program, and a specialized text editor.

Under the Polyglossum project, Lingvistica '93 and ETS Publishers Ltd. are developing two large general dictionaries to be distributed on compact disks later this year: German-Russian-German and French-Russian-French.

A German-Russian Bidirectional MT System

Lingvistica '93 has developed a new MT system in the 'PARS' series - PARS/Deutsch 1.0 - for translating between German and Russian. The system runs under Windows 3.1 and Windows 95. It can be started directly from MS Word 6.0 or MS Word 7.0, and it also translates HTML and Windows texts such as screen Helps.

We began developing a German to Russian MT system back in 1993, on the order of the Izvestia Concern, Moscow, Russia. It was supposed to run under DOS and produce draft translations of socio-political texts such as information messages by the VWD Agency. The prototype version was released in 1994, having a basic dictionary of 18,000 words and phrases.

In April 1997 work resume with financing from Igor Jourist Verlag, Germany. Within 5 months, a Windows bidirectional system was made based on the main PARS principles and having 3 dictionaries: general (23,000 words and phrases in each part, German-Russian and Russian-German); business (13,000); medical (8,000), made out of the corresponding Polyglossum dictionary compiled by ETS Publishers, Russia.

PARS/D shares the features of other PARS systems, together with various peculiarities due to specific features of German spelling and grammar.

For more information: Dr. Michael S. Blekhman, Director, Lingvistica '93 Co., 94a
Prospekt Gagarina, apt. 111, Kharkov 310140, Ukraine. (Tel.: +380 (0572) 27-71-35;
Email: blekhman@kpi.kharkov.ua or: blekhman@lotus.kpi.kharkov.ua)

==

CONFERENCE REPORTS

TMI-97 meets in Santa Fe
23-25 July 1997

John Hutchins

The magnificent surroundings of mountainous northern New Mexico was the invigorating setting for the Seventh International Conference on Theoretical and Methodological Issues in Machine Translation. The program combined the usual wide range of reports on current research activity with some retrospective reflections occasioned by the marking in 1997 of fifty years since Warren Weaver first suggested that computers might be used to translate natural languages.

In this spirit, the program included two speakers from the earliest period of machine translation and computational linguistics. Victor Yngve had been the head of the research project at MIT from 1953 until the mid 1960s. He told us about the directions of research at that time, his discussions with Noam Chomsky (then a member of the same research laboratory), and the development of the first non-numerical programming language COMIT. This was a prelude to a description of his more recent work on the theoretical foundations of a new approach to linguistics, based on truly scientific principles and not distorted by false assumptions going back to the earliest speculations by Greek philosophers. It was his belief that MT researchers could be more open to new ideas and methods than traditional linguistics, as demonstrated by their adoption of corpus-based approaches. Yngve's radical re-formulation of linguistics could not of course be expounded in the span of a relatively short conference paper, but we were urged and inspired to read his recently published book (*From grammar to science.* Amsterdam: John Benjamins, 1996).

The second speaker was George Miller, who had been involved with MT and information retrieval at Harvard University. His subsequent career had ranged over virtually the whole spectrum of computational linguistics and cognitive science, but recently he has become best known for the creation and development of WordNet. He devoted his talk to describing his current research on 'computer-assisted reading'. The basic idea is that children would learn to read from an electronic text where every substantive word would be linked to dictionary and encyclopedia entries. It would help children to understand any words they have difficulty with, and to build up their vocabulary. The idea has obvious wider applications, for students reading science textbooks, and for students learning foreign languages. The relevance of corpus-based MT research to the latter was obvious, and Miller encouraged participants to investigate this and other potential applications of their work.

One other speaker also informed the conference about earlier research. This was John Hutchins in a talk devoted to describing the beginnings of MT activity in the 1940s and in particular the events during and after the first MT conference in 1952. Otherwise, all speakers chose to speak about their own current and forthcoming research.

As might be expected, the host organisation (CRL, NMSU) was well represented. A joint paper by Kavi Mahesh, Sergei Nirenburg and Stephen Beale described word sense disambiguation using relationships in a domain knowledge base. The topic was taken up in a second paper by the same group presented by Stephen Beale.

Research for speech translation was a dominant topic. Shigeki Matsubara and Yasuyoshi Inagaki (Nagoya University) outlined a method of incremental parsing, transfer and generation taking English spoken input sequentially and producing Japanese output before sentence completion. Hideki Mima presented a paper from ATR on tackling problems of Japanese honorifics in spoken dialogue translation, applying situational knowledge in a basically transfer-driven MT model. There were two reports from the Verbmobil project: Birte Schmitz (Technical University Berlin) on the identification of turntaking markers in dialogue, and Jan W. Amtrup (University of Hamburg) on a unification-based architecture suitable for spontaneous speech.

The problem of 'natural' generation was tackled by Shiho Ogino and Tetsuya Nasukawa (IBM Tokyo), describing a method based on shallow discourse analysis; by Eduard Hovy and Laurie Gerber on using paragraph structure to improve Systran output; and by Karin Harbusch (University of Koblenz) who sought to define how discourse representation theory might be satisfactorily introduced into MT systems; and Antonio Sanfilippo (Sharp Laboratories, Oxford) suggested a thesaural approach to lexical selection.

The application of MT on the Web was the theme of Yumiko Yoshimura and colleagues from Toshiba, who described a method for Japanese Web browsers to identify English texts on sought subjects by means of proper nouns and statistical keyword analysis.

Francis Bond and colleagues described research for the NTT Japanese-English system in two papers: one on the analysis of temporal expressions (dates, time of day, weeks, months, etc.), and the other on the treatment of adverbial expressions. On similar practical lines was the paper from Kazunori Muraki and his NEC colleagues on lexical frames for selecting Japanese verbs

Efforts to improve MT quality were evident in many papers. Kanlaya Naruedomkul and Nick Cercone (University of Regina) described an iterative approach to quality improvement, using a combination of HPSG parser, selection constraints and word order patterns. The languages involved are Thai and English.

The control of input was the theme of Anna Sågvall Hein (Uppsala University), who described the Multra system for generating English and German translation for the Scania company from Swedish documents (some 6,000 pages in 1996). A checker for a controlled language 'ScaniaSwedish' has been developed based on the analysis component of the MT system.

Two papers from Spanish researchers were devoted to the finite-state neural-network model under investigation for the European Union-funded project EuTrans, basically an example-based system for Spanish and English. The claim is that 'subsequential transducers' are conceptually simple but computationally powerful finite-state models capable of incremental improvement from training data.

Other aspects of example-based approaches were discussed in papers by Ralf D. Brown (Carnegie Mellon University), on using statistical methods for dictionary extraction for the DIPLOMAT system; and by Brona Collins and Padraig Cunningham (Trinity College Dublin), on the use of 'adaptation knowledge' of previous successful matches. The more general problem of word alignment in bilingual corpora was tackled by Mathis Chen and others from the National Tsing Hua University, Taiwan, arguing for the use of 'topical' clustering of bilingual dictionary entries.

Arturo Trujillo (UMIST) described a method for improving chart generation in a lexicalist approach. The shake-and-bake approach was also the inspiration for the work of Fred Popowich and co-workers at Simon Fraser University and TCC Communications. The goal is a system for on-line translation of television captions and subtitles from English into Spanish (and later other languages). An obvious challenge is the treatment of colloquial and idiomatic language with high level of ambiguity, and within an unrestricted domain. On the other hand, recipients have the help of the visual images for understanding. It is further demonstration of the ever widening field of automatic translation to areas inconceivable to the pioneers of the 1940s and 1950s.

The final highlight of the conference was a 'dialogue' between Yorick Wilks and Sergei Nirenburg on "What's in a symbol? Ontology and the surface of language." Each put their different views on a series of issues, including, e.g. the nature of representations languages (are they natural languages or not? are they as ambiguous as natural language? are their primitives NL words or word senses?); their extensibility; the relation between NL and RL (is it that of language and metalanguage?); the differences between representation for human users/developers and representation for machines; and much more besides. Such was the depth and importance of the questions raised and the insights brought to them by both disputants that it is very much to be hoped that this dialogue is published in the near future for the benefit of everyone (and not just those fortunate to have been in Santa Fe in July 1997.)

TMI-97 was undoubtedly another successful conference in this continuing important series. It was ably and efficiently organised by the Computing Research Laboratory, New Mexico State University (Las Cruces) under the general chairmanship of Sergei Nirenburg. Harold Somers (UMIST) was the program chair, who is to be congratulated for bringing together an excellent group of speakers for the seventh of this series of successful biannual conferences.

The Spoken Language Translation Workshop
ACL/EACL-97, 11 July 1997, Madrid

Steven Krauwer, Doug Arnold, Walter Kasper,
Manny Rayner, Harold Somers

One of the many workshops organized in conjunction with the joint ACL/EACL Conference in Madrid this summer was dedicated to Spoken Language Translation (SLT).

The workshop was co-organized by ELSNET (the European Network in Language and Speech) and the editors of the journal Machine Translation. It attracted 40 participants and the programme included twelve presentations and three posters. Twelve out of fifteen contributions originated from the Far East and from Europe (six each), and three from the US. The workshop had four sections.

Exploiting and Exploring Dialogue Structure.

Dialogue MT introduces interesting problems beyond the already difficult issues of integrating speech processing with translation. As has clearly been recognised in the three papers which made up this opening session, identifying the special pragmatic features of spoken dialogue which distinguish such ``texts" from the type of input that a traditional MT system might deal with is a crucial part of the problem. Traditionally, the incorporation of contextual knowledge into an MT system was just dismissed as impractical, or at best, uneconomical. In a dialogue system, such an approach is unthinkable.

Manfred Stede and Birte Schmitz initiated the workshop with a close look at "discourse particles", the little words which can carry so much meaning, especially in terms of the overall dialogue structure. An additional problem is that many of these particles are ambiguous in that they also have an interpretation not related to discourse structure. Jae-won Lee looked at words whose translation is particularly dependent on the context, a problem which is exacerbated in a language-pair such as Korean--English. Their approach is to apply a statistical model of dialogue structure based on trigrams of speech acts. Keiko Horiguchi discussed meaningful ``errors" in speech which convey contextual meaning or the speaker's attitude, and then focused on the translation of discourse particles from Japanese into English. The approach adopted here is an analogical framework using a Cascaded Noisy Channel Model.

Dealing with Differences

The three presentations in this section explored some aspects of the SLT problem, which highlight the differences between translation of written and spoken text. Yumi Wakita described a method, which attempts to extract the parts of a spoken utterance, which have been reliably recognized, ignoring those which represent probable recognition errors. They presented results indicating that their method has an appreciable effect on the performance of a Japanese-English speech translation system.

Keiko Horiguchi and Alexander Franz described another piece of work aimed at counteracting the problems involved in taking translation input from a speech recognizer. They presented an example-based hybrid approach containing aspects of both corpus-based and rule-based styles of translation architecture; this move towards hybrid architectures seems to represent a strong tendency in current work within the field of SLT.

Finally, Pascale Fung presented a paper focussed on the problems which a speech recognizer has to contend with in a multilingual environment, where people typically speak using a variety of languages and accents. The talk described initial experiments which investigate the parameters of the problem, and in particular explored the possibility of constructing recognizers capable of recognizing multilingual input.

Towards Efficiency

The papers in this section addressed problems of efficiency in two senses. On the one hand SLT has to meet specific requirements of efficiency and robustness in processing, because speech recognition is imperfect, spoken utterances are often linguistically not well-formed, the translation must be available nearly simultaneously with the utterance, the quality of the translation must be sufficiently high as in most applications post-editing is not possible.

To solve these problems finite state transducer technologies are often employed and investigated. The papers by Alshawi and Amengual discussed different approaches along these lines. Both attempt to gain additional efficiency by a tight integration of analysis and transfer instead of assuming two different processing stages.

Another efficiency problem is that of acquiring the knowledge for building such a system. SLT systems are often heavily restricted to specific domains and in their vocabulary. This raises the question how such systems can be adapted to new domains and vocabulary. Therefore corpus-based statistical methods for language modelling and automatic acquisition are of special interest for SLT as addressed by Amengual and Frederking.

Methodological Issues

The common theme for the final three papers in the workshop was an emphasis on methodology and architecture. The first paper in the section, by Lavie, focused on the issues that arise when one transfers from a relatively narrow domain (in this case, Appointment Scheduling dialogues) to a broader domain (Travel Planning dialogues). The paper described some preliminary results of making this transfer for the JANUS system, and some modifications that may be required. In the second paper in this section (by Carter et al.), the main issue was not how one can broaden or enlarge the domain of a system, but how one can move from one domain to a distinct, potentially unrelated, domain of similar size. In other words, the focus was on the problems of customizing systems for new domains and languages. They argued that the characteristics of the Core Language Engine facilitate this customization.

The final contribution in this section was Mark Seligman's, which took a personal perspective in identifying six areas of SLT research as particularly interesting. (1) He argued the need for interactive disambiguation, and (2) for a particular kind of system architecture. (3) The third issue he addressed is that of how Speech Recognition and MT techniques should be integrated - in particular, whether a single set of techniques can or should be used to cover both tasks. He suggested that this is promising, though there are technical problems. (4) Seligman's fourth issue was how far natural pauses can be used in segmenting utterances, and how far analysis and translation can proceed on the basis of such segmentation. (5) The fifth issue recognized the importance of Speech Act identification in dialogue translation, and considers how a defensible and usable classification may be found. (6) Finally, there was the question of how one can restrict the range of candidate lexical items that have to be considered at each point in processing, and how candidates can be weighted appropriately.

Concluding Remarks

Some 15 years ago, when Machine Translation had become fashionable again in Europe, few people were prepared to consider seriously embarking upon SLT. After all, where neither machine translation of written text, nor speech understanding or speech production had led to any significant results yet, it seemed clear that putting three not even halfway understood systems together would be premature, and bound to fail.

Since then, the world has changed. If we look at the papers presented at the workshop we can clearly see that many researchers, both in academia and in industry, have taken up the challenge to build systems capable of translating spoken language.

This does not mean that anyone claims that most of the problems involved in speech-to-text, text-to-text translation, and text-to-speech have been solved. Although we have made tremendous progress, both from a scientific and from a technological point of view, many of the fundamental problems in MT and in speech understanding remain unsolved. Nevertheless, the movement towards more specialized systems, the redefinition of the notion of success, and the potential of dialogues, give us reason to believe that we will see many successful spoken translation systems in the future.

Recent Advances in Natural Language Processing (RANLP'97)

11-13 September 1997

Kalina Bontcheva
(University of Sheffield)

The second biennial conference RANLP was held again in Tzigov Chark, situated in the beautiful Rhodope mountains in Southern Bulgaria. Traditionally, the main focus was on recent developments, both theoretical and practical, in the field of natural language processing. The 58 papers (including the invited talks) covered many topics, techniques and applications - statistical tagging, word sense disambiguation, semantics, discourse, generation, machine translation, information extraction, text classification, language learning, lexicons, corpora, and tools. An introduction and overview of the state of the art on some of these topics were given by distinguished researchers at the summer school "Contemporary Topics in Computational Linguistics" which was held immediately before the conference (lecturers: Y.Wilks, S.Nirenburg, M.Zock, P.Seuren, T.McEnery, H.Trost, R.Mitkov, S.Sheremetyeva, M.Kudlek).

The three days of the conference were hardly enough for the presentation of all papers (3 invited talks, 28 regular, 12 short, and 5 reserve papers) which necessitated the organisation of parallel sessions for all short papers. An evidence for the high rating of the event was the number of submissions (more than 160) and also the increased number of participants and authors - about 70 people from 24 countries (17% from Spain, 13% from Germany, 10% from USA, 8.5% from Japan and UK, etc.)The invited speakers touched on important issues and even two of them presented alternative approaches to the problem of word sense disambiguation: Y.Wilks talked about "Combining independent knowledge sources for word sense disambiguation"; S.Nirenburg argued how full ontological knowledge (Mikrokosmos) can be used for word sense disambiguation; P.Seuren presented "A discourse-semantic account of topic and comment".

Many papers addressed various MT-related problems, e.g. word sense disambiguation (the two invited talks); semantic-based transfer (B.Wolf, M.Dorna, "Using hybrid methods and resources in semantic-based transfer"); information structure (P.Paggio "Information structure and MT: generating Danish existential sentences); shake-and-bake MT (D.Turcato et al "Inflectional information in transfer for lexicalist MT"); lexical ambiguity (B.Pedersen "Lexical ambiguity in MT: using frame semantics for expressing systemacies in polysemy"); example-based MT (T.Veale, A.Way "GAIJIN: a bootstrapping approach to example-based MT"); German-Russian MT (B.Staudinger, N.Smith "Aspect calculation in MIROSLAV: a German-Russian MT system"), and tools for building MT systems - Episteme (J.Amores, J.Quesada "Efficiency and elegance in NLP: the EPISTEME approach") and Amalia (S.Wintner et al. "Amalia, a unified platform for parsing and generation").

The majority of presented techniques were based on statistics and large corpora, which naturally resulted in an increased awareness for evaluation and efficiency. The implementation and testing of robust techniques based on a large-scale data seem to be gaining importance together with building re-usable language modules as needed for language engineering.

Apart from the formal presentations, there were several additional demonstrations of the implemented systems (e.g., a system presented by David Turcato). The increased attention towards working NLP systems might necessitate the allocation of dedicated system demonstration slot in the programme of the next conference.

The social calendar was also rich with events. It started with a traditional excursion to the old town of Plovdiv with its beautiful Renaissance architecture. In order to stimulate the contacts and informal discussions, the first day ended with a cocktail reception. The scenic area of the Batak lake was also a wonderful place to relax after the busy hours of the talks. The efficient organisation of transportation and accommodation before, during and after the conference, contributed to the pleasant stay of the attendees too.

Since the conference is held biennially, watch for announcements of the next event. Detailed information about RANLP'97 - titles of the papers, addresses of participants, etc. - is available from the conference web site at: http://www.cogs.susx.ac.uk/lab/nlp/ranlp/97.html.

The LISA Forum, Asia

Software Localization Professionals Focus on Emerging Asian Markets
Beijing, China, 6-8 August 1997

[Press release]

More than 150 localization professionals from all around the world gathered to discuss the challenges of the emerging Asian markets at the Localisation Industry Standards Associations annual Asian conference, hosted this year by WORD HOUSE, Hewlett Packard and the China State Bureau of Technical Supervision, in conjunction with the TSTT'97 International Conference on Terminology, Standardization and Technology Transfer.

Welcoming speeches were made by Jiao Yunqi, Director of CSICCI (the Chinese Standardization and Information Classifying and Coding Institute) and Alan Turley, Senior Trade Officer at the US Trade Department.

The Forum's opening keynote speech, "Telecommunications: Doing Business in China" was given by Henry Chang, General Manager of Nortel Post and Telecoms Inc., a Nortel Service Joint Venture headquartered in Beijing. In his speech, Henry Chang examined the risks, opportunities and mystique surrounding the Chinese market, and provided practical advice for foreigners wishing to succeed in it. "As the world's leading supplier of digital network solutions and services, Nortel is playing a strategic role in the development of China's telecommunications infrastructure, providing the latest technology and innovations for a networked digital economy. In terms of doing business in China, it is back to the basic 4 Ps plus guanxi (relationships) - a very important factor", said Mr. Chang.

In a complementary presentation, Christopher Chung, Business Manager, Microsoft IBN, gave an overview of doing business in Korea, which in his view will become one of the premier information industry markets by the beginning of the next decade.

Panel discussions and workgroups of experienced software/hardware manufacturers and services vendors looked at how to manage the Asian localization process, concentrating on the different markets and the business models and strategies required to deal with them, and on topics such as organizational support, staff skill sets, technical requirements, quality assurance, and cost-effectiveness.

Particular attention was paid throughout to cross-cultural issues, which are vital to the success of both the localization and management process and the finished products themselves. Supplementing these discussions were in-depth client/vendor case studies and operations reviews, which gave concrete examples of the issues discussed, and demonstrations of tools for localizing into Asian languages. Other, more specific issues covered were the Internet in China, multibyte enabling and leveraging simplified versus traditional Chinese, plus a pre-Forum workshop on copyright protection, organized jointly by LISA and Infoterm/TermNet.

Summing up, Michael Anobile, Director of LISA said that "the Beijing Forum demonstrated the increasing importance to the industry and maturity of the Asia-Pacific region, which is the fastest growing localization market by far. The in-depth discussions of topics such as Asian Web localization and Asian localization quality assurance underline LISA's role as the premier association and natural home for industry players throughout the world." All members and interested parties were invited to contribute actively to the Association's activities.

For more information: LISA Administration. 2 bis rue Ad-Fontanel. CH-1227 Carouge/Geneva, Switzerland. (Tel: +41 22 301 5760; Fax: +41 22 301 5761; E-mail: lisa@lisa.org; http://www.lisa.unige.ch

===

ASSOCIATION NEWS

AAMT has new Secretary General

AAMT has announced the retirement of Mr. Takashi Tanaka. His post of general secretary has been filled since August 1997 by Mr. Yasuo Nakajima. Previously Mr. Nakajima was at the Fujitsu Co. as manager of its Personnel and General Affairs Department.

European Association for Machine Translation

Minutes of the EAMT General Assembly held at the University of Copenhagen the 21st of May 1997 at 1630 hrs. (Minutes: Viggo Hansen, EAMT Secretary)

The General Assembly was attended by 24 members (Exhibit ‘A’) and called to order by the EAMT President, John Hutchins, who chaired the General Assembly.

1. Approval of Agenda

The Chairman reviewed the agenda following which the Assembly approved the agenda unanimously.

2. Executive Committee reports

President’s report

The President received the acceptance of the Assembly that agenda item 4 ‘Report on activities’ be part of the President’s report.

In his report the President mentioned the Vienna EAMT-workshop held in August 1996 (60 participants) and the 1997 EAMT-workshop in Copenhagen. The President expressed his appreciation to the organizers and the workshop-chairpersons, Dimitrios Theologitis and Bente Maegaard respectively. The President also reported on events having EAMT as joint sponsor, such ad ASLIP in London and two Evaluation workshops in Dublin.

Finally, the President informed about the new members campaign and the possibility of paying membership fee with credit card.

Secretary’s report

The Secretary (Viggo Hansen) informed the Assembly of the membership status. Before the EAMT workshop 1997 the membership was made up by 78 individual members, 6 non-profit making institutions and 9 commercial companies. In connection with the 1997 workshop about 20-25 individual new members were signed up giving a total of approx. 100 individual members.

The Secretary referred to the work done by Christine Favre at the permanent secretariat so generously hosted by ISSCO and Maghi King. The Secretary moved that the General Assembly expressed its appreciation to Maghi King and to Christine Favre for their assistance and work for EAMT. (Motion carried by acclamation.)

Treasurer’s report

The Treasurer (Doris Marty-Albisser) presented the 1995 and 1996 EAMT accounts. (Exhibit ‘B’ and ‘C’). The organization is in a financial healthy condition having assets totalling approx. 25000 Swiss Francs.

The Treasurer encouraged members to sign new members to the organisation and in particular new corporate members.

3. Newsletter report

The MTNI-editor (John Hutchins) gave a report on the newsletter situation. Certain Production problems had unfortunately caused late circulation of the newsletter to the members. Actions will be taken to try to speed up the distribution process.

The Editor announced his intention to retire as MTNI-editor.

4. Report on activities

This agenda item was partly dealt with under agenda item 2. President's report. The President added to his report information about the promotional brochure produced by Colin Brace. The President moved that the General Assembly expressed its appreciation to Colin Brace for his excellent work. (Motion carried by acclamation.)

5. WEB-pages

The creator of the EAMT Web pages, Colin Brace reported on EAMT's entry into the Internet.

6. Proposed amendments to the Articles of the Association

Reference is made to the proposed amendments as circulated to the members with the General Assembly convocation. The Secretary moved that Article 16, 17, 18, 19 and 20 be amended to read as following:

Article 16. Convening of the General Assembly

The General Assembly shall be convened by the President of the Association or, in case of incapacity of the latter, by the Executive Committee.

The ordinary meeting of the General Assembly must be convened once every year at such time as decided by the Executive Committee.

An extraordinary meeting of the General Assembly must be convened if members representing one fifth of the total voting strength of the Association's General Assembly requests so by registered letter addressed to the President of the Association.

The General Assembly shall meet in Europe. The venue for the meetings is decided by the Executive Committee.

Article 17. Procedure of convening the General Assembly

Both the ordinary and the extraordinary General Assemblies are convened by mail to each member and by notice published in MT News International, the official publication of International Association of Machine Translation at least four weeks before the date of the meeting.

The items on the agenda shall be mentioned in the convocation. Proposals to be considered by the General Assembly including proposals to in- or decrease the membership fee and proposals to amend the Articles and Bylaws of the Association shall be mentioned in or enclosed with the written convocation. They shall be held at the members' disposal at the Association's registered office.

Unless it is decided by a majority of two thirds of the votes cast by members present, no resolution may be adopted except those specified in the agenda apart from a resolution to convene an extraordinary General Assembly.

Article 18. Voting rights

As different classes of membership are reflected in the membership fees, so are they reflected in the voting rights:

- individual members are each entitled to one vote;
- members who are non-profit or profit making institutions are each entitled to two votes.

Article 19. Quorum and resolutions

The General Assembly is entitled to deliberate validly regardless of the number of members present.

The General Assembly shall adopt its resolutions and shall proceed with its elections by a majority of two thirds of the vote cast by the members present or represented by proxies bearing written powers. A resolution may be proposed by any member of the Association, but must be seconded by another member present at the General Assembly.

The resolutions of the General Assembly may also be taken by circular letter, by Fax or by electronic mail. Any resolution approved in writing by two thirds of the total voting strength by the Association's members shall be considered to have validly been taken by the General Assembly.

Article 20. Record and lists of attendance

The Executive Committee shall keep record including a list of attendance and the resolutions and elections of the General Assembly; it shall be signed by the President of the assembly as well as by its author.

Within thirty days following the assembly, the record shall be sent by the Executive Committee to the editor of MT News International for inclusion in said magazine.

Motion to amend the articles as proposed carried unanimously.

7. 1998 Membership fees

The Treasurer moved that the 1998 membership fees should be as following:

Individual member:	50 SFR
Non-profit making institution:	175 SFR
Company:	350 SFR

Motion carried.

From the floor it was proposed that a smaller fee should be maintained for students. The Secretary moved that the General Assembly authorize the Executive Committee to introduce a special low-fees-scheme for students and to specify conditions and rules for qualifying to the lower fee. (Motion carried unanimously.)

8. Election of officers

As the term of office is three years and all members of the Executive Committee were elected in 1995 no election took place. The Executive Committee members are:

John Hutchins, President
Viggo Hansen, Secretary
Doris Marty-Albisser, Treasurer
Colin Brace
Bente Maegaard
Dimitrios Theologitis
Jörg Schütz (MTNI)

9. Future activities

The President mentioned the following future activities:

- MT-Summit in San Diego, October 29 - November 1, 1997

- The 1998 EAMT-workshop. Venue and time not yet decided, but it is the intention of the Executive Committee to find a venue in the southern part of Europe. The European Association for Terminology has suggested a common arrangement between the two organizations, but no decision has been made as yet.

- MT-Summit 2001 will be in Europe. The President announced that bids for this event should reach EAMT by the end of 1997.

10. Any other business

Membership of EAMT normally qualifies for reduced seminar-fees to EAMT or EAMT-sponsored events. It was claimed that no reduced fee was obtainable at the Dublin-workshops. (The President undertook to investigate the matter.)

Seeing no other business the President adjourned the meeting of the 1997 General Assembly.

EAMT General Assembly EXHIBIT 'A'
21st of May, 1997

List of participants: Herman Caeyers, Robert Clark, Martha Ebermann, Hanne Fersøe, Viggo Hansen, John Hatley, Birgit Hoppe, John Hutchins, Iris Jahnke, Terence Lewis, Dirk Lueke, Hany Tawfik Kamel, Peter Kjeldsen, Monika Käser, Bente Maegaard (Proxy to Hanne Fersøe), Doris Marty-Albisser, Thorsten Mehnert, Dawn Murphy, Margrethe H. Møller, Roar Riseld, Joanna Stone, Dimitri Theologitis, Chris Thompson, Claire Trang (Proxy to Hanne Fersøe), Anne Tucker, Pia Wentzel

EXHIBIT 'B'

EAMT Statement 1995
(Adjusted statement as of December 31, 1995)

Income (CHF):
Membership fees

Individuals (net)	2918,00
Non-profit organizations (net)	700,00
Profit organizations (net)	2100,00

Non members (MT-news)	350,00
Total membership fees	6068,00
Miscellaneous (Yellow Book)	11,45
Special order MTNI (paid to EAMT)	107,45
Interest (net)	204,05
Total income	6379,50

Expenses	
EAMT contribution to IAMT (paid in ’96)	618,00
Newsletter (incl. 1. edition paid in ’96)	1240,45
Delivery charges UPS	23,10
Farewell gift MT Summit V	104,00
Bank Charges	175,50
Total expenses	2161,05

Net income over expenses 1995	4218,45
Total	6379,50
Balance of bank account in our favour	19418,25

Treasurer’s comment: Prepaid membership fees are booked in the financial year when payment is effected. The total of membership fees paid in one year thus provides the basis for the annual contribution paid to IAMT.

Pro memoria: Invoice no. 9510 (CHF 376.65) for the last ’95 newsletter edition was paid in January ’96. It is included in the 1995 EAMT statement. A special MTNI order (CHF. 96,00) from one member was paid directly to EAMT instead of IAMT.

EAMT Statement 1996
(Adjusted statement as of December 31, 1996)

Income (CHF):	
Membership fees	
Individuals (net)	2363,75
Non-profit organizations (net)	1225,00
Profit organizations (net)	3150,00
Non members (MT-news)	210,00
Total membership fees	6948,75
Interest (net)	204,05
Total income	7091,25

Expenses	
EAMT contribution to IAMT	703,00
EAMT brochure	841,25
Bank Charges	200,30
Total expenses	1744,55

Net income over expenses 1996	5346,70
Total	7091,25
Balance of bank account in our favour	23770,30

Treasurer’s comment: Prepaid membership fees are booked in the financial year when payment is effected. The total of membership fees paid in one year thus provides the basis for the annual contribution paid to IAMT.

Pro memoria: EAMT has not yet received any invoices from IAMT regarding the newsletter in 1996. Thus, appropriate provisions will be made in ’97. - By the time of closing the accounts, the results from TKE concerning their conference in Vienna, August 1996, with an EAMT workshop organized in conjunction with the conference, were not yet reported.

D. Marty-Albisser, General Assembly, Copenhagen 1997

NEWS OF RESEARCH ACTIVITIES

Projects at the Computing Research Laboratory, NMSU

Stephen Helmreich

The Computing Research Laboratory (CRL) at New Mexico State University, in Las Cruces, New Mexico, has a long history of academic research and development in multilingual text processing and related advanced computing applications. Several large-scale projects, such as Mikrokosmos, Corelli and OLEADA/Cfbola, along with a multitude of other research projects, have distinguished CRL's research and development efforts in machine translation.

Started in 1983, the lab employs 10 Ph.D.-level researchers and operates with an annual budget of more than $3 million. Although CRL is an academic lab, its central output includes a variety of working and deployed prototype systems for the applications studied--well beyond the publications usually expected from an institution of this kind. However, CRL researchers are publishing at a rate of more than 50 refereed journal and conference proceedings articles a year. Additionally, CRL has also been quite active in organizing national and international conferences and workshops. This past July, CRL sponsored the 7th International Conference on Theoretical and Methodological Issues in Machine Translation held in Santa Fe, New Mexico.

The R&D teams at CRL are actively pursuing research in most of the current machine translation (MT) paradigms: knowledge-based machine translation (KBMT), example-based machine translation (EBMT), transfer-based machine translation, multi-engine machine translation and statistics-based machine translation. MT configurations at CRL include fully automatic MT environments as well as environments for human-assisted MT and machine-assisted human translation. The languages CRL currently concentrates upon in its MT efforts include: Spanish, Arabic, Chinese, Japanese, Russian, Serbo-Croatian, Korean, Persian and English. In connection with its work on MT, CRL pursues research in computational morphology; syntax; ontological semantics, including development of large computational ontologies and interlingual text meaning representation languages; computational semantic analysis, including treatment of non-literal language (e.g., metonymies, metaphors, etc.); text planning and text generation; storage, search and control architectures; and user and knowledge acquirer interfaces.

CRL is the home of the Mikrokosmos, a large-scale knowledge-based, interlingual MT project led by Director Sergei Nirenburg. The system performs Spanish to English analysis and is being expanded to include Chinese. An English generation component is to be added later this year. At the core of the system is a 5,000-concept, 80,000-statement language-neutral world model, or ontology (developed by Kavi Mahesh, now at Oracle Corp.) which provides the language for meaning explication for entries in a Spanish lexicon of approximately 40,000 entries. The latter was constructed by manually seeding about 7,000 entries and then obtaining the rest automatically through the application of lexical rules. Evelyne Viegas led the lexicon acquisition effort, assisted by Victor Raskin. Stephen

Beale developed a general control and planning architecture known as Hunter-Gatherer, used in both analysis and generation in Mikrokosmos. This architecture is based on the combination of the branch-and-bound, constraint satisfaction and solution synthesis techniques and is extremely fast and convenient for natural language processing applications.

The CRL Corelli project, under the direction of Remi Zajac, is pursuing development of transfer-oriented MT between Spanish, Russian, Japanese, Russian, Serbo-Croatian and English. The first version of the system uses a document handling approach developed for the TIPSTER project, an ongoing effort among researchers and developers in government, industry and academia to create an integrated information retrieval and information extraction system. A newer version is CORBA (Common Resource Broker Architecture) compliant. The results of the TIPSTER project are being deployed by the intelligence community. The
TIPSTER project is funded by the Defense Advanced Research Projects Agency (DARPA).

The immediate goal is to configure an environment that allows for quick addition of new languages at a basic level that includes morphological analysis, lexicon and glossary lookup and a shallow translation. Further plans call for the inclusion of more involved syntactic analysis. Scientific advances in this project include a novel approach to morphological analysis developed by Svetlana Sheremetyeva, implemented by Wanying Jin and tested on Russian and Serbo-Croatian. The basic document and process architecture developed by Zajac is another innovation in this project, whose general approach is based on earlier efforts led by Nirenburg both at CRL and, earlier, at Carnegie-Mellon University.

Machine-Assisted Human Translation at CRL is pursued in the framework of the OLEADA/Cfbola project, led by William Ogden. The main goal of this long-term, multi-faceted effort is to provide multilingual text processing technology to language instructors, learners, translators and analysts. Starting with close observation of working habits of professional translators and constantly incorporating feedback from these users, the project now includes sophisticated alignment algorithms to provide a translation memory for translators, concordances, glossaries, morphological analysis and dictionary lookup. The system is currently being used by language instructors at government language schools to analyze online corpora for teaching examples.

Four new projects are beginning at CRL this fall. The first is the Expedition Project, an ambitious three-year project to build a system that can be used by a computer specialist and a language expert to produce a machine translation system within six months. This system will be including syntactic analysis, word sense disambiguation and coreference resolution in a new language.

A second project now beginning includes the development of Persian machine translation capability. The project will encompass a system that incorporates advanced Persian processing tools. Users will be able to consult, modify and enrich the system's dictionaries through specially designed user interfaces for processing new textual material.

A third project, MINDS (multilingual interactive document summarization), involves summarization, extraction and translation of documents in Japanese, Spanish and Russian. The MINDS project is focused on the creation of a multilingual summarization tool designed to provide quick and interactive document filtering, even in the absence of certain lexical or other resources for a language. The fourth project is Unicode Retrieval System Architecture (URSA) whose focus it is to make detection, retrieval and collection visualization transparent to query and document language issues. Ongoing work involves prototypes of cross-language information retrieval systems, the development of Unicode information retrieval technologies and close integration with the TIPSTER document management architecture.

While these projects comprise a major portion of CRL's focus, many other research and development projects are currently underway at CRL. They include:

* Multilingual information retrieval and language-based approaches to information retrieval in English, multilingual information extraction, text summarization, coreference resolution and proper name recognition and classification, led by CRL Deputy Director James Cowie.
* Pragmatics-oriented machine translation, pursued by David Farwell and Stephen Helmreich.
* Unicode and various other character encoding and font set capabilities, developed by Mark Leisher, who serves as a member of the Unicode Technical Committee.
* Authoring systems, with an initial concentration on authoring patent claims, developed by Svetlana Sheremetyeva. The first implementation of such a workstation, called ClaimWright, has been recently completed.
* Foundations of ontological semantics, a general approach to computational treatment of meaning in natural language, formulated by Sergei Nirenburg, along with other CRL personnel, as well as Victor Raskin of Purdue University and Yorick Wilks of University of Sheffield, a former CRL director.
* Reasoning about mental states that are described in natural language discourse with a special sub-focus on mental states that are described metaphorically in discourse, pursued by John Barnden of Birmingham University, UK, and Stephen Helmreich of CRL.
* Comprehensive collaborative planning environments for teams of human and computer agents, specifically, for the tasks associated with producing summary reports from a variety of multilingual information sources, including those on the Internet, researched by Sergei Nirenburg, James Cowie, Stephen Beale and Rémi Zajac.
* Computational studies of discourse, led by Janyce Wiebe of the NMSU Computer Science Department. Wiebe, together with Rebecca Bruce, of Southern Methodist University in Dallas, is pursuing work on corpus-based word sense disambiguation.
* User-centered interactive system design and multilingual information retrieval, researched by William Ogden.

Additionally, a variety of massive data acquisition tasks carried out at CRL by most of the aforementioned. They also train the acquirers, design the various acquisition toolkits and manage the actual acquisition work.

The CRL teams benefit from the experience of senior programmers, such as Ron Zacharsky, Yevgeny Ludovik and Nigel Sharples, senior knowledge acquirers, such as Lori Wilson and Jeffrey Longwell as well as a small army of junior research assistants, many of whom are NMSU graduate students.

CRL has close ties with other New Mexico State University (NMSU) departments, central among them are Computer Science, Psychology and Languages and Linguistics. Outside NMSU, CRL has cooperated (among others) with colleagues from Brandeis University, Carnegie-Mellon University, the Information Sciences Institute at the University of Southern California, New York University, University of Pennsylvania, University of Maryland and Southern Methodist University.

Funding sources over the years have included the Department of Defense, Defense Advanced Research Projects Agency, National Science Foundation, National Institute for Standards and Technology, Air Force Office of Scientific Research, Office of Naval Research, Internal Revenue Service, Lockheed, IBM, Apple Computer and various other private businesses.

Further information about CRL, its projects and researchers, can be found at the web site: www/crl.nmsu.edu/Home.html. Or, you may contact Linda Fresques, technical writer, at 505-646-6429.

The MULTIDOC Project:

MULTI-lingual product DOCumentation

Jörg Schütz, IAI Saarbrücken

1. Introduction

MULTIDOC is a European project of the Fourth Framework Programme within the Language Engineering Sector. It is founded on the specific needs and requirements of product documentation expressed by five representatives of the European automotive industry (Bertone, BMW, Renault, Rolls-Royce Motor Cars and Volvo) with particular focus on the multilingual aspects of product documentation. The general goal is to define and specify methods, tools and work-flows supporting stronger demands on quality, consistency and clarity in the technical information, and shorter lead times and reduced costs in the whole production value cycle of documentation including the translation into multiple languages. The results of the project are applicable to any other component or system manufacturing business; thus, they are not restricted to the automotive industry. The project is divided into two phases: an inception and elaboration phase, the so-called MULTIDOC Concerted Action (LE3-4230), and a construction or development phase, the so-called MULTIDOC Project (LE4-8323). The first phase has been finished recently and will be the main theme of this article. The second phase is to be started within the next months (December 1997).

2. Basic Requirements and Vision

The aim of the MULTIDOC Concerted Action was to identify the problem areas and to specify solutions for the European automotive industry when it comes to multilingual product documentation and also set a roadmap for the future. In software engineering, this phase is usually called the inception phase of an iterative software development process. During inception we establish the business rationale for the project and decide on the scope of the project. This is also the phase where we get the commitment from the project sponsor(s) to go further; in our case this was the successful evaluation of our MULTIDOC Project proposal.

The Concerted Action also included parts of the elaboration phase of a software development project. In elaboration, we collect more detailed requirements, do high-level analysis and design to establish a baseline architecture, and create the plan for construction which is the actual software production phase consisting of many iterations. In our domain, the most crucial bottlenecks comprise the following business areas that needed further elaboration:

* More and more languages in which product documentation has to be published; there is a drastically increased focus on Asian and East-European markets.

* Increasing costs for translations.

* Lead times in the document production process and in the translation process.

* Poor or no possibility to measure and control the translation process.

* Inconsistent use of information structure and information content.

All project partners agree that besides the quality of the product the services associated with the product and the accompanying documentation of the product must be seen as an integral part of the product. To satisfy the demand for high-quality technical documentation, the documentation has not only to be comprehensible and up to date, it has to be produced and delivered (including the accessibility to new or up-dated information) with modern technologies. The necessity of integrating services, documentation and networked information technology (IT) solutions with the support of modern, multilingual language technology (LT) forms the basis for the MULTIDOC vision of an Abstract Documentation Factory (ADF).

3. MULTIDOC Virtual Application

3.1. Translation Engineering

Within the Concerted Action a so-called virtual application was defined. It constitutes a compromise between the present situation of product documentation in the different automotive companies and the MULTIDOC vision of an ADF that is based on the concept of Translation Engineering. The strategy is based on the present situation and has to be maintained with various restrictions for the different automotive companies but with the common interest to work toward the ADF vision that is shared by all companies, however, with different ways to reach the vision. The virtual application is the result of the elaboration phase, and it allows for a smooth and cost effective transition of the business, because we have first and foremost concentrated on the existing process stages, where several control capabilities for the source language, such as spell, grammar and style checking functionality as well as the control of terminology consistency, support the technical writer and other knowledge workers in identifying and defining information objects in an SGML authoring environment. The virtual application already includes steps toward Translation Engineering -- the operational foundation of the MULTIDOC vision. Translation Engineering (TE) will revolutionise the current way of thinking in technical documentation because the whole documentation process is oriented toward multilingualism. This new business scenario includes a push/pull policy for technical information delivery and retrieval in an automotive dealer's workshop in combination with a translation-on-demand policy. TE is responsive to the new business demands, and it will harmonise and unify the most crucial documentation requirements in areas such as the consistent use of technical information in structure and content, the efficient and effective reuse of information objects based on standardised information structures, and the terminological and multilingual orientation of the whole information production process.

3.2. Abstract Documentation Factory

The vision of the ADF comprises the complete re-organisation of the documentation processes. This means that LT in general, and specifically multilingual LT including translation technology, will make the move from a supporting technology to an enabling technology. The focus of the ADF is on the following three main components:

Multilingual terminological ontology as a means for representing domain knowledge (the subject of technical documentation) linked with natural language semantics.

Object Modelling Technique (OMT) as a theoretical foundation for analysis and design, and as an implementation platform based on, for example, CORBA.

Agent technology as the overall umbrella for construction, and as an alternative implementation platform, especially for networked applications.

In the ADF, knowledge producer and knowledge consumer will operate in virtual environments brokered by software agents. A software agent acts autonomously on behalf of a person to fulfil the person's goal or task. Agents are also key enablers for push technology, which is used in information update tasks and information retrieval tasks.

3.3. Validation

All development strategies have been validated with a cost/benefit appraisal based on a hypothetical business calculation of a virtual automotive enterprise. We have taken this way to further maintain the generalisation direction, which we already followed in the other phases of the Concerted Action. However, our profitability assessment is based on actual calculations made by the MULTIDOC partners for their specific enterprise situation.

4. Conclusions and Perspectives

Both approaches, “bridging the gap” and ADF, are centred around the MULTIDOC terminological ontology as the primary information source. The parallel development allows for an optimal use of resources, and permits a straightforward implementation of the ADF based on already existing LT modules and components.

More details on the project can be obtained from our web pages at URL http://www.iai.uni-sb.de/MULTIDOC.

TAO group (Montreal) change address

Please note that the TAO group has moved to the University of Montreal. It is now called the RALI and the new address is: RALI, Departement d'informatique et de recherche operationnelle, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Quebec H3C 3J7, Canada.

OSCAR: Translation Memory Exchange Format Standards Initiative

[LISA Press release]

The formation of a new special interest group, OSCAR (Open Standards for Container/Content Allowing Re-use), was proposed at the June 1997 Forum held by the Localisation Industry Standards Association (LISA) in Washington DC. Aimed at discussing ways to standardize data exchange between various translation tools, the initiative was the result of a very successful meeting of leading tool developers and some of their clients held immediately before the Forum.

The ad hoc meeting held in Alexandria, Virginia, on June 2, 1997 was hosted by tools vendor Trados, chaired by Microsoft, and stimulated by the OpenTag proposal made earlier in the year by service provider ILE. Alan K. Melby of Brigham Young University was selected as Technical Secretary. The participants were drawn from major developers and users of translation tools and included AlpNet, IBM, ILE, ITP, Logos, Microsoft, Multiling, Star, Systran, and Trados.

The group agreed to cooperate under the umbrella of the Localisation Standards Industry Association in the development of an industry standard translation memory exchange format (TMX). Once the format is defined, each developer will be able to write a routine that will export to and import from it. Users can then export translation memory databases to this intermediate format and import the exchange file to another translation memory tool. They will be thus be able in the medium term to port memories between different tools, thus speeding up their work and protecting their own and their clients' investments in both tools and content.

A decision was taken that the initial standard should provide a high-level format dealing only with how segments of text are aligned, without specifying segment internals. However, a subsequent standard will address segment-internal aspects of the OpenTag format and other approaches to processing the markup codes inside text segments.

Information: LISA Administration, 2 bis rue Ad-Fontanel, CH-1227 Carouge/Geneva, Switzerland (Tel: +41 22 301 5760; Fax: +41 22 301 5761; Email: lisa@lisa.org; Web: http://www.lisa.unige.ch) Contact: Deborah Fry, 100637.711@ compuserve.com

==

CORPORA AND SERVICES

European Language Resources Association

ELRA announces update of its catalogue of Language resources for Language Engineering and Research. It currently consists of: (1) Spoken resources: 39 databases in several languages (recordings from microphone, telephone, continuous speech, isolated words, phonetic dictionaries, etc.); (2) Written resources: 14 monolingual and multilingual corpora, 28 monolingual lexica, around 60 multilingual lexica, a linguistic software platform and grammars development platform; (3) Terminological resources: over 360 databases with a wide range of domains and several languages (Catalan, Danish, English, French, German, Italian, Latin, Polish, Portuguese, Spanish, Turkish). Information: ELRA/ELDA, 87 Avenue d'Italie, 75013 PARIS (Tel: +33-1-45-86-53-00; Fax: +33-1-45-86-44-88; Email: info-elra@calva.net; WWW:
http://www.icp.grenet.fr/ELRA/home.html)

Parallel corpora

1. Parallel texts (English-French, English-Spanish) from the World Health Organisation (http://www-pll.who.ch/programmes/pll/cat/cat_resources.html)

2. Searchable Canadian Hansard French-English parallel texts (1986-1993) from Laboratoire de Recherche Appliquée en Linguistique Informatique, Université de Montréal. (http://www-rali.iro.umontreal.ca/TransSearch/TS-simple-uen.cgi)

CANCODE

Cambridge University Press and The University of Nottingham have for some time been engaged in compiling CANCODE - a corpus of naturally occurring speech in English. We plan to expand the corpus by several million words in the near future, and would like to come into contact with serious researchers who have - or plan to make - tape recordings of appropriate data.

For further details, contact: Jean Hudson, Research Editor, Cambridge University Press & University of Nottingham Spoken English Corpus Project (email: jhudson@cup.cam.ac.uk; Tel: +44-1223-325123)

Linguistic Data Consortium

CALLHOME Collection

The objective of the CALLHOME project is the creation of a multi-lingual speech corpus that will support the development of Large Vocabulary Conversational Speech Recognition (LVCSR) technology. The collection covers six languages, American English, Egyptian Arabic, German, Japanese, Mandarin Chinese, and Spanish.

SWITCHBOARD-1 Release 2

The Switchboard-1 Telephone Speech Corpus was originally collected by Texas Instruments in 1990-1, under DARPA sponsorship. The first release of the corpus was published by NIST and distributed by the LDC in 1992-3. SWITCHBOARD is a collection of about 2400 two-sided telephone conversations among 543 speakers (302 male, 241 female) from all areas of the United States. A computer-driven "robot operator" system handled the calls, giving the caller appropriate recorded prompts, selecting and dialing another person (the callee) to take part in a conversation, introducing a topic for discussion, and recording the speech from the two subjects into separate channels until the conversation was finished.

The Kids Corpus

This database is comprised of sentences read aloud by children. It was originally designed in order to create a training set of children's speech for the SPHINX II automatic speech recognizer for its use in the LISTEN project at Carnegie Mellon University.

CALLFRIEND Collection

The CALLFRIEND project supports the development of language identification technology. Calls were collected in the following languages: American English, Canadian French, Egyptian Arabic, Farsi, German, Hindi, Japanese, Korean, Mandarin, Spanish, Tamil, and Vietnamese. Two major dialect groups were collected for English, Mandarin, and Spanish. The dialect comparison groups include: southern vs. non-southern American English, Caribbean Spanish vs. non-Caribbean Spanish, and Mainland Mandarin (China) vs. Mandarin as spoken in Taiwan.

Boston University Radio Speech Corpus

The Boston University Radio Speech Corpus was collected by Mari Ostendorf of Boston University, primarily to support research in text-to-speech synthesis, particularly generation of prosodic

patterns. The corpus consists of professionally read radio news data, including speech and accompanying annotations, suitable for speech and language research.

DSO Corpus of Sense-Tagged English Nouns and Verbs

This corpus contains sense-tagged word occurrences for 121 nouns and 70 verbs which are among the most frequently occurring and ambiguous words in English. These occurrences are provided in about 192,800 sentences taken from the Brown corpus and the Wall Street Journal, and have been hand tagged by students at the Linguistics Program of the National University of Singapore. WordNet 1.5 sense definitions of these nouns and verbs were used to identify a word sense for each occurrence of each word. In addition to providing the word occurrences in their full sentential context, the corpus includes complete listings of the WordNet 1.5 sense definitions used in the tagging.

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

Southeast Asian Languages List (SEALANG-L)

[From Linguist List]

SEALANG-L is a non-moderated mailing list devoted to scholarly discussion relevant to Southeast Asian languages. While much past discussion has focused on 'traditional' linguistics, I am strongly encouraging participation from computational linguists and computer scientists. Research across the board has suffered greatly from a lack of instrumentation (on the linguistics side), analysis (on the software side), and machine-usable data (on both sides). Existing software/algorithms for both roman-alphabet and Far Eastern languages have not supplied satisfactory solutions. For more information, visit the SEALANG Web site: http://seasrc.th.net/sealang.

===

Publications Announced and Received

Machine Translation and Translation Theory

Edited by Christa Hauenschild and Susanne Heizmann. Berlin: Mouton de Gruyter, 1997. xiv, 266pp. ISBN: 3-11-015486-2. Price: DM.168,00; US$ 105.00

The volume contains papers given at the 2nd International Workshop 'Machine Translation and Translation Theory', held in 1994 at the University of Hildesheim, as well as invited contributions.

Contents: *Barbara Moser-Mercer*: Process models in simultaneous interpretation; *Hans G. Hoenig*: Using text mappings in teaching consecutive interpreting; *Christiane Nord*: The importance of functional markers in (human) translation; *Heidrun Gerzymisch-Arbogast*: Translating cultural specifics; *Monika Doherty*: Textual garden paths; *Birgit Apfelbaum and Cecilia Wadensjoe*: How does a Verbmobil affect conversation?; *Birte Prahl and Susanne Petzolt*: Translation problems and translation strategies involved in human and machine translation; *Susanne Jekat*: Automatic interpreting of dialogue acts; *Louis des Tombe*: Compensation; *Peter E. Pause*: Interlingual strategies in translation; *Birte Schmitz*: The translation objective in automatic dialogue interpreting; *Jan W.Amtrup*: Perspectives for incremental MT with charts; *Susann LuperFoy*: Discourse processing for voice-to-voice machine translation; *Margaret King*: Evaluating translation.

Language Today

Language Today is a new magazine for the world's language industries produced by Praetorius Ltd. in collaboration with the Logos Group of Modena, Italy. Its on-line edition can be found at http://www.logos.it/language_today. From September 1997 it will be published monthly: single issue £5.50, annual subscription £55.00. For further information: Language Today, Praetorius Publications, Suite 2b, Joseph's Well, Park Lane, Leeds LS3 1AB, UK. (Tel: +44 113 242 2255; Fax: +44 113 244 2965; Email: development@praetorius.com)

Journal of Computer Speech and Language

Special issue on Evaluation in Speech and Language Technology.

Papers on the topic of Evaluation in Speech and Language Technology are invited for inclusion in a special issue of "Computer Speech and Language" to be published in April 1998. This call is primarily directed at the authors of papers submitted to the SALT Workshop on Evaluation in Speech and Language Technology, held June 17-18, 1997, in Sheffield. However, other authors working this area are also encouraged to contribute papers. The editor for this special issue will be Dr. Robert Gaizauskas, Department of Computer Science, Sheffield University. Deadline for submission of manuscript: November 21, 1997. (Email: R.Gaizauskas@dcs.shef.ac.uk; Tel: +44 (0)114 222 1827; Fax: +44 (0)114 222 1810)

Language and Computation

Announcing a new journal edited by Dov Gabbay (Imperial College), Ruth Kempson (SOAS), Shalom Lappin (SOAS), and Uwe Reyle (Stuttgart). Language and Computation is an independent electronic and paper journal devoted to the publication of high level research papers on issues in the interface of logic, linguistics, formal grammar, and computational linguistics. It will be published quarterly, is sponsored by Oxford University Press and FOLLI, and distributed freely from a web site at Imperial College, London. The articles will be made available as compressed ps files which can be downloaded from the web site. We hope to bring out our first issue early in 1998. At the end of our first year of publication, we will consider the possibility of distributing a hard copy of the first volume of the journal (distributed by Oxford University Press) to libraries and subscribers. Executive Editor: Hans Juergen Ohlbach, Department of Computing Science, Imperial College, London, UK, ho1@doc.ic.ac.uk.

Human Language Technologies: Living and Working Together in the Information Society

Publication of Discussion Document, Luxembourg, July 1997. The document is the intermediate result of an ongoing concerted effort in defining future European RTD activities in Human Language Technologies, started in May 1996 and involving a large number of actors from industry, research laboratories and academia.

The idea at the root of the document is that RTD in Human Language Technologies should be focused on a small number of challenges, central to key drivers of the Information Society, and to which it can contribute in a substantial way. It should build on existing European strengths and potentialities, address the European RTD policy issues, namely the criteria set out for the Fifth Framework Programme, meet the aspiration of European citizens and support a cohesive development of the Information Society of the 21st century in the EU.

The work described in the document, in particular in Part III, is an implementation of this model. It is intended to be both focused and sufficiently open-ended for a flexible implementation, reflecting the reality of changing priorities. Specific RTD options will be the subject of further consultations with users, developers and researchers. Besides the RTD activities proper, substantial support for focused demonstration actions and basic infrastructure is foreseen.

The document is available on the web at: http://www2.echo.lu/langeng/en/fp5/lt.html; or by contacting: LINGLINK, Anite Systems, 151 rue des Muguets, L-2167 Luxembourg (Tel: +352 427744; Fax: +352 439594, Email: linglink@anite-systems.lu)

PUBLICATIONS RECEIVED

Journals

AAMT Journal *no.19, June 1997, no.20, September 1997.* In Japanese only.

Computational Linguistics *vol.23 no.1 (March 1997).*

ELRA Newsletter *vol.2 no.1 (March 1997).*

Elsnews *vol.6 no.2 (1997).*

Language International *vol.9 no.3 (June 1997)*. Contents include: Confessions of a software localizer (Yann Meersseman). -- Future tense for Euro Systran? (Colin Brace). -- Focus on Czech translation tools; *vol.9 no.4 (August 1997)*. Contents include: Boom times for Euro Babel (Andrew Joscelyne). -- Hurry up and wait: the cost of localizing too early (Yann Meerseman). -- Teletranslation comes of age (Minako O'Hagan). -- Langenscheidt's T1: old wine in new bottles? (Deborah Fry). -- Machine translation at the crossroads (John Freivalds)

LISA Newsletter *vol.6 no.3 (August 1997)*. Contents include: Building empires in quicksand (Yann Meerseman). -- Rising to the challenges of Asian translation and localization (Minako O'Hagan). -- Software conversion for the Chinese market (Lloyd Yam).

Literary and Linguistic Computing *vol.12 no.3 (September 1997)*

Localisation Ireland *vol.1 issue 3 (September 1997)*. Contents include: TRADOS leads the way.

Machine Translation *vol.12 no.1-2 (1997) Special issue: New tools for human translators*, ed. Pierre Isabelle and Kenneth W.Church. Contents include: The proper place of men and machines in language translation (Martin Kay). -- MT today: emerging roles, new successes (Mary Flanagan). -- Bricks and skeletons: some ideas for the near future of MAHT (Jean-Marc Langé et al.) -- A technical word- and term-transition aid (Pascale Fung and Kathleen McKeown). -- Termight (Ido Dagan and Ken Church). -- Multilingual document production (Anthony Hartley and Cécile Paris). -- Glossary-based MT engines (Rémi Zajac and Michelle Vanni). -- Accessing foreign languages with COMPASS (Elisabeth Breidt and Helmut Feldweg). -- Target-text mediated interactive machine translation (George Foster et al.); *vol.12 no.3 (1997)* Contents: From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology (John Hutchins). -- Book reviews.

Multilingual Computing *vol.8 no.3 (1997)*. Contents include: Web translation made easy (Scott Nesbitt). -- Russian word processing with Windows 95 (Galina Raff and Peter Cassetta). -- How translation tools help (Willem Stoeller). -- HTML authoring for translation (Ultan Ó Broin); *vol.8 no.4 (1997).* Contents include: The language market through 2005 (Joel Sawyer). -- Defining a translation database exchange format (Deborah Fry). -- International software quality assurance (Pauline Cho)

Books

The LISA Directory. Localisation Industry Standards Association, June 1997.

Reports

Human language technologies: living and working together in the information society. Discussion document, Luxembourg, July 1997.

Language engineering: progress & prospects. Luxembourg: LINGLINK, [1997]. 118pp. Published on behalf of the Language Engineering Sector of the Telematics Applications programme, European Union.

Conference proceedings

Spoken language translation. Proceedings of a Workshop sponsored by the Association of Computational Linguistics and the European Network in Language and Speech (ELSNET), 11 July 1997, Universidad Nacional de Educación a Distancia, Madrid, Spain. Edited by Steve Krauwer, Doug Arnold, Walter Kasper, Manny Rayner, Harold Somers. [ACL, 1997] viii, 97pp. [Copies available from ACL, P.O.Box 6090, Somerset, NJ 08875, USA. Email: acl@bellcore.com; http://www.aclweb.org]

TMI 97. Proceedings of the 7th International Conference on Theoretical and Methodological Issues in Machine Translation, July 23-25, 1997, St.John's College, Santa Fe, New Mexico, USA. [Las Cruces: CRL, 1997] iv,224pp. [Report of conference elsewhere in this issue.]

Items for inclusion in the 'Publications Received' section should be sent to the John Hutchins 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.