

# elsnews

The Newsletter of the European Network in Language and Speech

Feb 1998

# 7.1

## The Multilingual Information Society

Jan Roukens, European Commission

The phrase *Multilingual Information Society (MLIS)* was probably used for the first time in 1994 as the title of a draft *Communication and Community* programme of the European Commission. The concept emerged from an environment that focused on information and telecommunication technology, including language engineering applications. In conjunction with the Information Society initiative, it marks a major paradigm shift: the object of the new programme is not technological development itself, but the linguistic needs of a society evolving rapidly under the influence of technological development and internationalisation in all spheres of human activity.

The MLIS concept resolves a number of factional debates of the past, mostly territorial struggles, as to what are language technologies and what not, what are language industries, and how language technologies relate to "applications" and other technology-driven undertakings. These questions appear a bit irrelevant now. The heated debates about the closeness of speech processing to language processing cooled down when speech translation came into view.

MLIS helps to dissolve artificial walls. In doing so, it broadens the scope for a richer and more diversified scenario of options for research, technological and linguistic development, socio-economic initiatives and language policy debates. What really matters is whether an initiative contributes to resolving language (or communication) problems. This can be approached from many points of view and different positions on the upstream-downstream curve. We are allowed to think today of an information society in which people continue to use their own languages, in the understanding that everyone will make an effort to communicate across language borders as well. In the future, any tool, service or system will need to accommodate or be adaptable to any language. The European Community will focus on European languages, in the framework of standards and agreements, taking into account this global scope.

The need to adapt to the language of the citizen, customer or client, and of various language markets, has repercussions both for private companies and public institutions in European countries and regions, and for the Community as a whole.

It has also re-opened the millennia-old discussion on whether the number of languages should be less, or the number of interlinguas limited. Esperanto and Latin have reappeared on the scene as politically neutral alternatives to English, French, German, Spanish and, who knows, Mandarin-Chinese.

There will never be one single language regime for all circumstances, of course. Looking ahead to the next century, it seems likely that English will remain the language used internationally for science and global business for some time to come, even if national scientific exchanges will continue to be conducted in the national languages. At the same time, entertainment TV for the general public will continue to be dubbed or subtitled, and one cannot imagine that companies would conduct electronic commercial transactions with customers in another than the customer's language.

In the 21st century people will work and live in a highly translated world. They will translate in their own minds, and they will have translations made for them. Quality of translation will be an important determining factor for the quality of life. Economically speaking, translation will be big business, allowing ideas and facts generated anywhere on the globe to be spread everywhere. Any method or tool that will reduce the effort involved in this while improving the quality of translation will be welcomed by the industry.

While in the foreseeable future people in Europe will continue to speak a large number of languages, questions are raised about this trend continuing in the very long term, and whether the evolution should be steered towards more pluralism, or less. The MLIS programme clearly maintains the status quo in this respect, by encouraging the various language communities to take measures that allow and facilitate the use of their languages in the electronic environment.

It would be wise, however, to start socio-economic research around this theme, to underpin the political choices that will continue to be made and to guide investment decisions in the private and public sectors. And, last but not least, to help the citizens who have to weigh the personal investments made in their mother tongues against the cost of a re-orientation in this multilingual world.

Multilingual Document Management	2
Donia Scott & Roger Evans	
Controlled Languages	4
Pim van der Eijk	
New multilingual applications from Sharp	6
Ian Johnsom	
MATE Project update	7
New Directions for MLIS	8
Interview with Susan Armstrong	
LISA The Localisation Industry Standards Association	9
SpeechMania	10
Gerrit Bloothoof	
ELSNET Summer School Announcement	11
Future events	11

Feb 1998

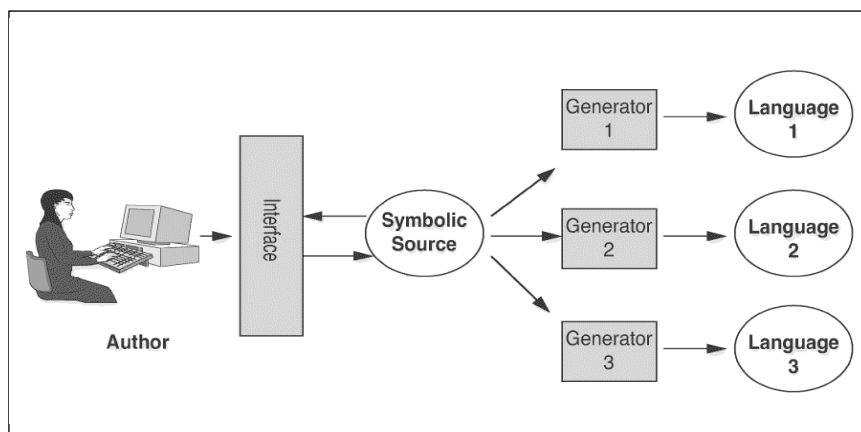
elsnet  
.....

# Multilingual Document Management Without Translation

## Using Natural Language Generation in the Multilingual Information Society

*Donia Scott & Roger Evans, Information Technology Research Institute, University of Brighton*

One of the core activities underpinning the information society is document management — the ability to create, maintain and update documents, and especially sets of related documents, in a coordinated way. Multilingual document management, where documents need to be maintained in several languages, will similarly be the foundation of an effective multilingual information society. Current approaches to multilingual document management employ manual or automatic translation (or a combination of



*Symbolic authoring architecture for multilingual document production*

both) from master monolingual sources. Recent developments in Natural Language Generation (NLG) technology suggest an alternative approach in which the master source is language-neutral, and documents in all languages are generated independently and automatically. This eliminates 'source language bias', makes subsequent updates to a document easier, faster and probably cheaper, and facilitates multilingual maintenance of the document base.

Technological support for monolingual document management is now quite well-established. Template and stylesheet facilities are common on many word processors, access and version control supports co-ordinated document development, and macros and conditional constructions can be used to support different variants of the same basic document. But the multilingual situation brings with it additional problems. The most fundamental is how to maintain 'over time' versions of the same document in different languages. The techniques for management of variants of a document in the same language are in general not powerful enough to support the relationship between the same document in different languages, even when they are quite direct translations of each other. To make matters worse, in general direct translations are not what is required: different languages and cultures have their own ways of expressing the same ideas, and the most effective document is one which conforms in style as well as language to the reader's expectations. Supporting this requires techniques far beyond the abilities of most current document management systems.

At present, the principal way of producing versions of a single document in several languages is through translation: the document is initially written in one language and then translated into other desired languages. Manual translation is big business, but it is costly (good translators are relatively rare and therefore expensive) and always takes place under

time pressure. Automatic translation is potentially quicker and cheaper, but current systems still lack the quality, coverage and adaptability required to deliver final copy of important public documents.

In addition, translation-based multilingual document management tends to favour the source language. The appropriate style, register and distance from the reader for a particular document type varies from language to language, as does the linguistic realisation of these features. For example, in instructional texts, French is more likely to use indirect constructions than English, and also more likely to express them using impersonal pronouns rather than passive constructions (Paris and Scott 1994). Expert translators (with no time constraints!) can accommodate these differences, but more often echoes of the source language detract from the quality of the translated document.

The alternative approach that we and our colleagues have been exploring uses a technique called **Symbolic Authoring** to generate language-neutral symbolic representations of the content of a document, from which documents in each target language are generated automatically, using NLG technology. NLG has been developing steadily in recent years, and a number of commercial or near commercial systems now exist<sup>1</sup>. Many of these systems take their input from some external data source. The idea of Symbolic Authoring is simply to allow the user to specify the generator input directly.

<sup>1</sup> For example **AlethGen** (Coch 1996), **CORECT** (Levine et al, 1996), **DRAFTER** (Paris et al, 1995), **EXCLASS** (Caldwell and Korelsky, 1994), **FOG** (Goldberg et al 1994), **GhostWriter** (Marchant et al, 1996), **GIST** (Power et al, 1995), **IDAS** (Reiter et al, 1995) and **ILEX** (Knott et al, 1996), **LFS** (Iordanskaja et al 1992), **ModelExplainer** (Lavoie et al 1996), **PlanDoc** (McKeown et al 1994) and **PostGrapher** (Fasciano and Lapalme 1996).

ELSNNews is published at the Centre for Cognitive Science, University of Edinburgh using Aldus Pagemaker™. It is printed on recycled paper by Lutton Press, Ltd.

Editors: Mimo Caenepeel, Mariken Broekhoven and Steven Krauwer.

Lay-out and production: Mimo Caenepeel

ISSN 1350-990X

© ELSNET 1998.

FOR INFORMATION Contributions to ELSNNews, and address corrections, should be sent to:

elsnews@let.ruu.nl

Tel: +44 131 650 4594

Fax: +44 131 650 6626

Material for the next issue is due:

15 March 1998

e

In essence, a Symbolic Authoring system comprises a natural language generator coupled to an interface that supports the manual creation of the generator's input (that is, the authoring of the symbolic (conceptual) content of the document). Such a system becomes interesting if we add additional generators for other languages (as in the figure on the opposite page). Now a single (symbolic) authoring process supports multilingual variants of a document directly: one update to the document is reflected in all languages simultaneously. Furthermore, each generator can be tuned to its own language and cultural settings, choosing its own most appropriate realisation strategy independently of the others.

As well as the NLG technology, it is clear that the other key requirement of a Symbolic Authoring system is an effective user interface. The 'symbolic content' required by an NLG system is typically a LOOM-like knowledge base (MacGregor, 1988), and the user interface must enable the author to construct such a knowledge base. This is a significant problem, which different systems have addressed in different ways. Our own most recent work uses a technique called WYSIWYM ('what you see is what you meant' (Power et al. 1997)) to present the knowledge base to the author as text (using the same NLG technology as the authoring component itself). Early experiments suggest this could be a very effective and general solution to the input interface problem.

Symbolic Authoring allows the simultaneous management of a document in several languages, through the use of a language-neutral content representation. These 'symbolic sources' can themselves be managed as documents (sharing structure, using macros and templates and so on). The symbolic nature of the information also allows for more powerful authoring support such as cross-referencing, consistency checking and stylistic control. Additionally, because the source documents are language-neutral, they can be maintained equally well by authors of any nationality (using appropriately localised interface tools — and with WYSIWYM, this localisation comes for free). The authoring language is purely a feature of the interface, not the underlying document.

How much of what we have described is feasible right now? Current NLG works best with fairly short documents in well-understood genres (such as instructional texts). In addition, existing input representations tend to be quite application-specific. Nevertheless, systems such as DRAFTER, GIST and GhostWriter show that useful applications can be created within those constraints. Effective symbolic authoring user interfaces exist, and there are exciting developments in this area, such as WYSIWYM. Full integration into a real document management system also remains an outstanding task, but a primarily technical one. In summary, most of the key pieces of this potential cornerstone of MLIS are there, just waiting to be put together.

## References

- Caldwell, D. and Korelsky, T. (1994): *Bilingual generation of job descriptions from quasi-conceptual forms*. In: Proceedings of the Fourth Conference on Applied Natural Language Processing.
- Coch, J. (1996): *Evaluating and comparing three text production techniques*. In: Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING-96).

Fasciano, M. and Lapalme, G. (1996): *PostGrappe: a system for the generation of statistical graphics and text*. In: Proceedings of the Eighth International Workshop on Natural Language Generation Herstmonceux, Sussex, UK.

Goldberg, E., Driedger, N. and Kittredge, R. (1994): *Using Natural language Processing to Produce Weather Forecasts*. IEEE Expert, 9 (2).

Iordanskaja, L., Kim, M., Kittredge, R., Lavoie, B. and Polguere, A. (1992): *Generation of Extended Bilingual Statistical Reports*. In: Proceedings of the fourteenth International Conference on Computational Linguistics (COLING-92).

Knott, A., Mellish, C., Oberlander, J. and O'Donnell, M. (1996): *Sources of Flexibility in Dynamic Hypertext Generation*. In: Proceedings of the Eighth International Workshop on Natural Language Generation, Herstmonceux, Sussex, UK.

Lavoie, B., Rambow, O. and Reiter, E. (1996): *The Model Explainer*. Demonstration presented at the Eighth International Workshop on Natural Language Generation, Herstmonceux, Sussex, UK.

Levine, J., Rogers, I., Bennington, T. and Pattison, C. (1996): *Class hierarchies as a Multi-Purpose Knowledge Representation in a Requirements Capture and Design Tool*. In: Proceedings of Expert Systems'96, Cambridge, UK.

MacGregor, R. (1988): *A Deductive Pattern Matcher*. In: Proceedings of the 1988 Conference on Artificial Intelligence. St. Paul, Mn, USA.

Marchant, B.P., Cerbah, F. and Mellish, C. (1996): *The GhostWriter Project: a demonstration of the use of AI techniques in the production of technical publications*. In: Proceedings of Expert Systems'96, Cambridge, UK.

McKeown, K., Kukich, K. and Shaw J. (1994): *Practical Issues in Automatic Documentation Generation*. In: Proceedings of the Applied Natural Language Processing Conference, Stuttgart, Germany.

Paris, C. and Scott, D. (1994): *Stylistic Variation in Multilingual Instructions*. In: Proceedings of the Seventh International Workshop on Natural Language Generation, Kennebunkport, Maine, USA.

Paris, C., Vander Linden, K., Fischer, M., Hartley, A., Pemberton, L., Power, R. and Scott, D. (1995): *A Support Tool for Writing Multilingual Instructions*. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Canada.

Power, R., Cavallotto, N. and Pemberton, L. (1995): *The GIST Specification Tool*, LRE Project 062-09, Deliverable PR-3b.

Power, R., Scott, D. and Evans, R. (1997): *What You See is What You Mean: direct knowledge editing with natural language feedback*. Technical report ITRI-97-03, Information Technology Research Institute, University of Brighton, UK.

Reiter, E., Mellish, C. and Levine, J. (1995): *Automatic Generation of Technical Documentation*. Applied Artificial Intelligence, vol 9, no 3.

## FOR INFORMATION

Donia Scott and Roger Evans can be contacted at:  
 Information Technology Research Institute  
 University of Brighton,  
 Lewes Rd,  
 Brighton BN2 4GJ, UK  
 Tel: +44 1273 642 900  
 Fax: +44 1273 642 908  
 Email: {Donia.Scott, Roger.Evans}@itri.brighton.ac.uk  
 WWW: <http://www.itri.brighton.ac.uk>

Feb 1998

**elsnet**  
 .....  
 .....

# Controlled Languages in Technical Documentation

*Pim van der Eijk, Cap Gemini ATS*

*Controlled Languages (CLs) are precisely defined subsets or variants of language constructed for use in particular environments for specific purposes. They are used chiefly for the authoring of technical documentation, and to support translation.*

*CLs have been around for a while: the first controlled language for technical documentation, *Caterpillar Fundamental English (CFE)*, was developed in the 1960s. But the concept has recently seen a vast increase in popularity: these days, it is adopted in some form or other by hundreds of companies and organisations. Pim van der Eijk explains that this is due to increased quality requirements and economic globalization, and describes how CLs contribute to the quality and comprehensibility of technical documentation, and reduce the amount of time spent on (and the costs involved in) translation.*

## CLs and Technical Documentation

When technicians perform operations, maintenance procedures and fault diagnostics on complex technical systems, they need recourse to technical documentation on the tasks and procedures involved. The quality of such documentation is obviously critical: if it is inaccurate, incomplete or difficult to understand, this can affect the system's downtime, result in

insufficient command of formal technical English to interpret complex procedural information correctly. And, for reasons of cost and time-to-market, translation of the documentation may not be an option.

Figure 1 gives an example of the text type and application domain of technical documentation: it shows an Interactive Electronic Technical Manual (IETM) for a generator developed for a defence application, and illustrates the complexity of technical information conveyed.

## CLs and Translation

From the very beginning, CLs have been associated with Machine Translation (MT), or with translation in general. Given the tiny fraction of the world's annual translation volume that is currently performed using MT, it should be clear that the most obvious benefit that CLs bring to technical translation is that comprehensibility of source documentation is as beneficial to human translators as to any other reader. It prevents misinterpretations (which obviously may have dramatic consequences) and can actually help translators make significant time savings.

Major vendors of MT software strongly encourage use of CL input, as it reduces the overall MT post-editing effort. This is most pronounced when the MT system's dictionaries are "tuned" to the source CL. Leading vendors of products for terminology management and Translation Memory (TM) report similar benefits. These systems are productivity tools for human translators that operate by finding fuzzy matches and generating proposal translations based on previous translations and terminological data. The lexical and syntactic standardization a CL brings to source documentation increases the hit rate of these systems, and thus overall human translation efficiency.

Experience in CL implementation has shown that many domains are inherently too complex to allow for the drastic reduction or elimination of ambiguity that some CLs originally aimed at. For instance, in the *Caterpillar* domain there are seven senses for the term

*valve*, and complete elimination of this ambiguity is just not feasible. A common solution adopted in many CL projects is to obtain disambiguation information from authors, and to store this information with the source data using an SGML encoding. In subsequent stages this encoding can be used by the IETM system for glossary generation and hyperlink generation, or by an MT or TM application for target language vocabulary selection.

This type of disambiguation encoding demonstrates the pragmatic approach that is increasingly being adopted in implementations of CLs. The CL, and its support environment, is just one factor in the entire documentation production chain, in which judicious use of SGML encoding, TM, terminology management, MT and other mechanisms all combine for optimization of the full documentation process.

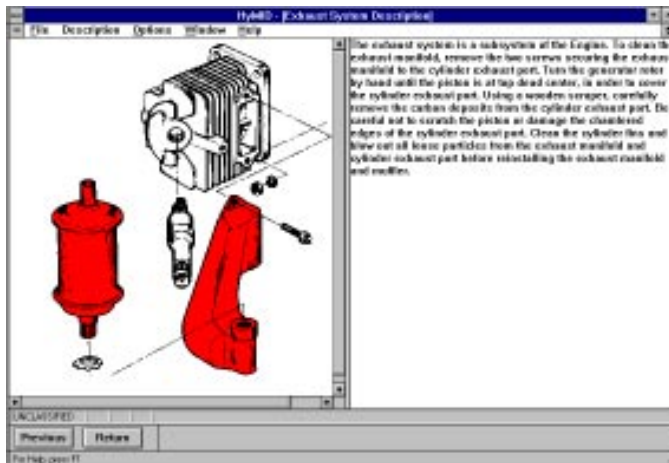


Figure 1: Interactive Electronic Technical Manual

incorrect parts replacements or, more seriously, cause damage to expensive equipment and injury to humans. The latter may in turn lead to costly liability claims against the supplier. This is one area where CLs are playing an increasingly important role: the use of CLs in the authoring of technical documentation improves its quality and comprehensibility.

Another reason for the increasing use of Controlled Languages in technical documentation is economic globalization: companies in many sectors are increasingly selling their products to a global, and therefore multilingual customer base. In the case of high-tech products, such as medical equipment, military aircraft, or computer software, vendors often find that more than half of the intended target users have an

e

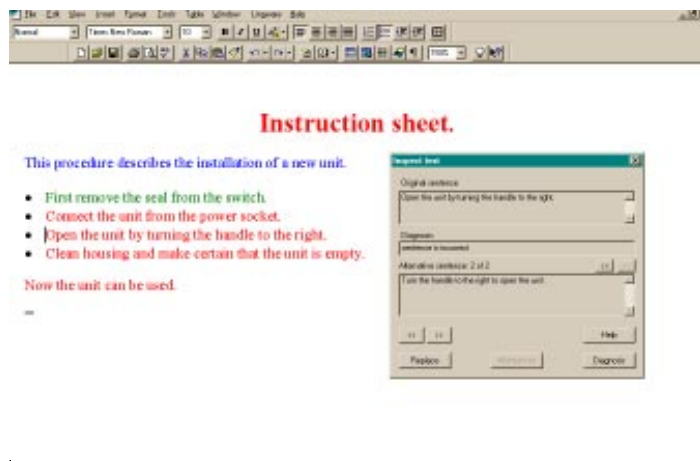


Fig 2 Interactive CL Correction.

## Controlled Language Authoring Support

Several large companies have developed CL support tools using in-house staff. There are also a small number of companies in the computer services industry, including Cap Gemini, that offer CL support tools on a per-project basis. These tools in general aim to meet all or a subset of the following requirements:

- Linguistic analysis of Controlled Language text.
- Generation of useful *critiques* to authors.
- General morpho-syntactic and spelling correction.
- Support for interactive *transformation* of general sublanguage expressions into the Controlled Language.
- Integration in standard DTP environments.

From the requirement of interactive transformation it follows that a correction system has to model uncontrolled author input and its mapping to correct CL expressions. As this mapping is generally many-to-many, an interactive system can offer the user a list of alternatives (preferably ordered using some plausibility metric) from which (s)he can choose. *Figure 2* shows a screen example of such a correction session using an integration of Cap Gemini's SE correction software with Microsoft Word.

## Who uses CLs?

Nowadays, the concept of CL is adopted in some form or other by hundreds of companies and organisations. It ranges from the use of (often rather informal) company-internal guidelines for technical writing, and lists of preferred and unallowed terminology, to professional SGML-based document authoring systems that use full parsing to enforce validation of an application-specific CL grammar. Apart from English, there are controlled variants of French, German and Swedish.

However, most applications of CL are unknown outside the site at which they are used. There are three reasons for this. First of all, technical documentation, in particular at the level of lexicon and terminology, is inherently domain-specific, or even specific to a particular company, type of product, and target user. This reduces the portability of a CL to other domains. Furthermore, many companies that develop and use Controlled Languages see no need to disclose to outsiders what, to them, represents proprietary knowledge and experi-

ence, and may in some cases offer them a competitive edge. And finally, the general NLP research community seems to view the technical communication domain as a somewhat uninteresting application area, rather than recognizing it as the rich source of research challenges it is to those who are more familiar with it.

There are some exceptions to the limited familiarity of CL. The best-known of these is the AECMA **Simplified English (SE)** standard, developed for aerospace documentation, which is by far the most well-researched controlled language in practical use. SE has been researched extensively, and has been shown to meet the objective of improved comprehensibility in a statistically significant way, and in situations

where it is most relevant (namely in complex procedures that are to be performed by non-native aircraft maintenance technicians). Given the safety-critical nature of procedural information in aerospace documentation, and the considerable cost of aerospace equipment, these results are encouraging and exciting: they show that compliance to SE, and investment in SE implementations within an organisation, are not just a matter of meeting industry standards, but are an investment for which an independent business case can be made.

## Conclusion

With the growing complexity of industrial systems and increased globalization, readability and comprehensibility of technical documentation will only become more important in future. Simplified English presents a successful instance of a CL that meets these objectives, and will reinforce interest in CL and the need for CL authoring support in industry. Language control also contributes indirectly to the overall productivity in a full documentation lifecycle, by reducing human translation time, improving the hit rate of TM systems, and reducing post-edit efforts when using MT systems. In general, there is an increasing tendency towards pragmatic solutions, where CLs are one ingredient of a broader organizational and technical framework.

In the meantime, CL designers and specification committees would very much benefit from research by linguists and communication researchers to help them determine precisely *which* features of a CL contribute *most* to comprehensibility, and at which *cost* in terms of authoring complexity.

### FOR INFORMATION

Information on Controlled Languages and NLP related to Controlled Languages is available on the Internet at <http://www-uilots.let.ruu.nl/Controlled-languages/>. This page has pointers to several conferences, mailing lists, and has a bibliography which includes on-line readable versions of some articles on the subject.

**Pim van der Eijk** can be reached at  
Cap Gemini Nederland BV  
PO Box 2575  
3500 GN Utrecht  
The Netherlands  
Tel: +31 30 252 6240  
Fax: +31 30 252 7045  
Email: [pvdeijk@inetgate.capgemini.nl](mailto:pvdeijk@inetgate.capgemini.nl)

Feb 1998

**elsnet**  
.....

# The Multilingual Information Society: New Applications from Sharp

*Ian Johnson, Sharp Laboratories of Europe Ltd.*

Sharp Laboratories of Europe (SLE), based in Oxford, was founded in 1990 as Sharp's first R&D laboratory outside Japan. Sharp's global strategy towards R&D has since led to the establishment of a further laboratory in the USA. A principal motivation for this globalization of the R&D process was a recognition of the increasing importance of developing products which addressed the multilingual needs of people around the world. The advent of the World Wide Web, and the ability of people to access more and more information in a diversity of languages and in various media, has had a major impact on the demand for multilingual applications of language technology.

Sharp's interest in language technology products dates back to the early 1980s, when the DUET E/J family of translation products was started. The first product, sold in 1987, was based on a Unix workstation and included an OCR. The cost of the system was in the region of £20k, and the target market was professional translators and translation agencies. The system has been refined over the years — in particular it has been integrated with a Web browser — and the latest Power E/J system is now available, with Sharp's PC or as a stand-alone software package selling at around £60. The increasing affordability of PCs has opened up the market for these translation products. And the possibilities for accessing documents on the Internet has led to an increased demand for on-line translation software, not just for professional translators but also for people who would simply like to access information in other languages.

Sharp is committed to improving global information access and communication for people living in the Multilingual Information Society. Two recently developed personal translation products are representative of this drive: the **Multilingual Document Generator**, and the **Sharp Intelligent Dictionary**.

The **Multilingual Document Generator (MDG)** was the first product of the *Language & Information Systems Group* at SLE. Its objective is to enable users to

generate documents in a target language they have little or no knowledge of. The application allows them to customise standard documents written in their own language, and to generate translations of these documents in other languages with guaranteed linguistic, stylistic and cultural correctness. By providing people with the ability to communicate in the language of the recipient, it enables them to overcome language barriers encountered either in their business or in personal correspondence. As the user would not normally be able to assess the quality of the text generated, the MDG output has to be 100% correct. The latest UK version of the system generates documents prepared in English into French, German, Spanish and Italian. The product recently won an award in the New Technologies section of the 'Languages for Export 97' scheme organised by the *London Chamber of Commerce and Industry*.

The process of implementing this first version of the MDG has given us many ideas and insights into ways in which such systems could be improved. We have carried out a detailed analysis of the existing system and produced an initial design document for a second version. Major improvements will include extension to further languages (including Japanese); increasing the number of document templates provided (particularly the more business-oriented ones); provision for more flexible command types to permit greater flexibility between languages; access to on-line bilingual dictionaries; and porting the system to other hardware platforms (e.g. PC and PDA).

Our second product, the **Sharp Intelligent Dictionary (SID)** (English-Japanese version), has recently been released on the Shoin word processor and will shortly appear on Sharp's Multimedia Notebook PC. Its aim is to allow people to browse documents on the Internet in a language they may not be familiar with.

SID does not carry out full translation, but looks up words and phrases in the context of the sentence in which they are found. The major benefits for the user

over existing systems are that SID provides fast, accurate lookup of words and phrases (including discontinuous collocations) in a dictionary. The system works out the most likely translations for the words and phrases making up the sentence, and then presents them to the user. It is primarily aimed at people looking at English documents on the Web who have some knowledge of English but need help in understanding particular words and phrases.

It is interesting to note that Sharp (and other vendors of MT systems), realising the difficulty of achieving accurate MT particularly with long sentences, have recently been offering translations of phrasal units as an alternative to full translation. SID is perhaps best thought of as a browser or a glosser, allowing the user to scan information quickly and independently before deciding whether to download the document and possibly have it translated in full. The system can also be used as an aid to professional translators who could use its output to assist them in producing their own translations. The figure on the opposite page shows a screenshot of the SID user interface. Notice in particular the underlining which marks collocations and menus for browsing alternative translations.

In providing these multilingual applications, Sharp is tackling the problem of computerised translation of documents from two perspectives. In the case of the MDG there is an absolute requirement to ensure that the output document is 100% correct, because it is assumed that users themselves will not be able to judge the accuracy of the translations generated. In the case of the SID, on the other hand, the aim is not to achieve 100% correct translation, but to allow users to combine their own knowledge of the language with the information provided by the system, to gain an understanding

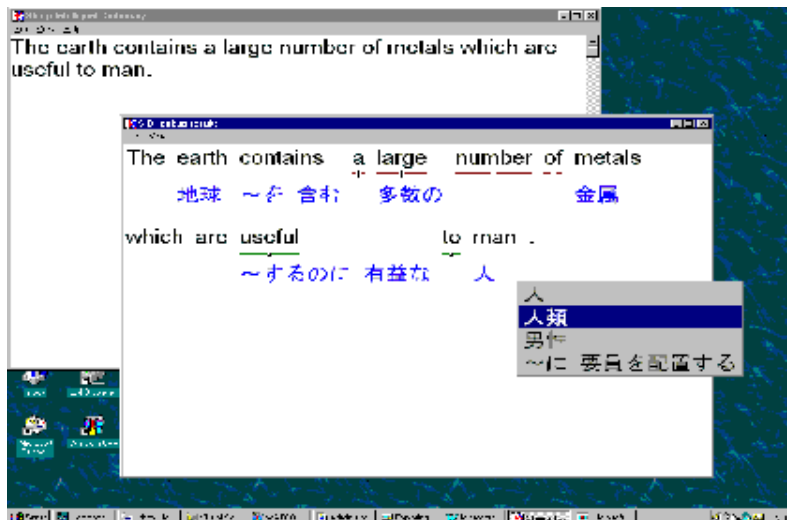
<sup>1</sup> As English is the dominant language on the Web, Sharp are focussing on translation applications from English into other languages. Our first language pair is English-Japanese, but further pairs are planned.





of the text concerned. With translation into the user's own language we can rely to a large extent on the user being able to piece together the words and phrases presented by the system into a coherent whole.

Sharp is committed to providing a range of personal translation tools, ranging from document generation to lookup of words and phrases in a variety of dictionaries, in order to help users approach foreign languages with greater confidence. The degree of linguistic competence expected of users ranges from zero (in the case of the MDG) to some basic knowledge of English (in the case of SID). Since the MDG provides templates which are guaranteed to be both linguistically and stylistically correct. However, it also caters for people with more advanced knowledge of a foreign language, by helping them determine, for instance, the correct stylistic way of expressing oneself (for example in the case of a letter of condolence). Similarly, SID may be useful even to those with an advanced knowledge of English, because the system allows the selection of different user levels, thus ensuring that advanced users are



*A screenshot of the SID user interface*

not shown translations for relatively simple words. The system also has a variety of technical dictionaries which have been developed in the course of these R&D activities.

By providing personal translation products such as those described here, Sharp is carrying out its mission to improve global information access and communication in the Multilingual Information Society.

#### FOR INFORMATION

Ian Johnson can be contacted at:  
Sharp Laboratories of Europe LTD  
Edmund Halley Road,  
Oxford Science Park  
Oxford OX4 4GA  
Tel: +44 1865 747711  
Fax: +44 1865 714170  
Email: ianj@sharp.co.uk

## MATE: Multi-level Annotation Tools Engineering

The market for products which are linguistically enhanced is potentially enormous but currently underdeveloped. One of the root causes of this is the high cost of producing language resources of sufficient scope and quality for everyday applications. Language engineering projects typically either develop the resources they need themselves, or acquire resources from the results of previous projects and adapt them specifically to novel purposes. As a result there is a wide variety of annotation schemes, and a few tools to handle annotated dialogues, but there are no standards or general methodological guidelines for the creation, annotation, retrieval and analysis of annotated dialogue corpora. This situation makes it hard for any spoken language dialogue system developer to reuse tools in different projects.

The MATE project will review results from projects across the world and use them as background for the provision of annotation guidelines, i.e. a preliminary form of standard. It will also develop a workbench based on portable, open standards (such as WWW and Java), which will enable partners and prospective users to collaborate even from a distance. Corpora provided by the partners will be used for testing and evaluation.

The outcome of MATE will be a set of standards for the annotation of dialogue corpora, covering the levels of prosody, (morpho-)syntax, co-reference, dialogue acts, and communication aspects, as well as their interaction. A work-

bench of tools, methods, and guidelines will support the use of the standards in creating, acquiring, maintaining, enhancing, and applying corpora.

#### FOR INFORMATION

The MATE project is funded by the Telematics Language Engineering programme, and will run for 20 months. Its expected starting date is 1 March 1998. ELSNET funded a pilot action in **Dialogue Annotation**. The results of this can be found in a paper by Hans Dybkjær and Ulrich Heid, on

<http://www.mip.ou.dk/nis/publications/papers/elsnet-da-96/elsnet-da-96.html>.

For information concerning MATE, please contact:  
Laila Dybkjær  
Maersk Mc-Kinney Moller Institute for Production Technology  
Forskerparken 10  
DK-5230 Odense M, Denmark  
Tel: +45 65 57 35 53  
Fax: +45 63 15 73 08  
Email: laila@mip.ou.dk  
<http://www.mip.ou.dk/nis/>

Project

Feb 1998

**elsnet**  
.....

# Old Customs and New Directions in the Multilingual Information Society

*What are the roles of academia and industry in facing the challenges of MLIS? Is there scope for closer collaboration between the two communities in this area? Could localisation profit from a more theoretical perspective on multilingual document handling? What is the perspective on the current upsurge of interest in MLIS in a country like Switzerland, where life in four languages has long been a everyday reality? And where have the recent big European projects on corpus-based research taken us — do we need a radically different approach to this kind of work?*

*Mimo Caenepeel talked about these issues to Susan Armstrong, Professor at the Ecole de traduction et d'interprétation at the University of Geneva and researcher at ISSCO (Istituto Dalle Molle per gli Studi Semantici e Cognitivi), a research lab attached to the University of Geneva. For the past ten years Susan has worked on various aspects of multilingual information processing, from machine translation and bilingual lexicography to corpus-based studies of texts and their translations. She has also been active in a number of initiatives for the acquisition and preparation of lexical and textual resources for NLP research and development. Her research has focused on methods and tools to exploit multilingual resources for translation studies, with special attention to the combination of linguistic and statistical methods.*



**ELSNNews:** You have worked in, and moved between, both academic and industrial working environments. How did you experience the difference between these environments? Is there profitable interaction between them, or could this be improved?

**Armstrong:** I have found that there is quite a gap between the two communities. The academic focus on teaching, research, projects and contributions to scientific organizations is quite different from daily business activities and long term commercial interests. However, on the practical level, there are many cases of shared interests as witnessed in the numerous collaborative language engineering projects. I'm also sure that many colleagues, like myself, have very good individual contacts with industrial colleagues.

One important area where interaction could be improved is in continuing education. There is a real need for the professional community to understand the potential of new technological developments. For multilingual NLP applications, in particular, the field is expanding rapidly in areas such as translation support tools, multilingual document management, and the use of translation memory. Both the technology and the products continue to evolve. How do professionals get an appreciation — and understanding — of the kind of technology that would support their work? In my view, the academic community is best situated to make this type of knowledge available to the professional public. It has the expertise and the most up-to-date information, and it has less vested interest than the commercial sector. But this is not something we have concentrated on so far; it has mostly been left to sellers of software, who, understandably, offer PR talk.

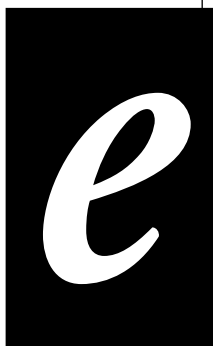
Technology transfer has been a popular term for some time now. We should extend this to include the end users of our technology. Finding the right level of information is important in this respect. I am currently teaching a course on technology for translators. These are people who do not have a background in linguistics or computer science, but who nevertheless need to be educated users, who need to know about the potential of new technologies, and how it fits into their day-to-day work. This is

even more important for the professional translator. In my experience many people working in industry are looking for more knowledge and understanding of technological solutions for multilingual language tasks. Communicating the relevant, often highly specialized, knowledge and expertise to interested professionals is tremendously important, and people in the academic community are well-placed to do so - but it is a real challenge.

**ELSNNews:** Localisation has a reputation for being “a-theoretical” in orientation. Some people claim it could benefit from (tools based on) a more theoretical perspective on multilingual language handling. What are your views on this?

**Armstrong:** Can you tell me of good theoretical work in this area that has practical applications, and is of real benefit to the relevant community? Take theoretical MT work, for instance — there has been no real progress for a while now in this area. Commercially useful systems are still working in the paradigms of the 70's. We have reached a point where only small improvements are possible, but nothing substantial. The fact is that people are always looking for new, and better, tools. This is a topic that comes up again at every LISA (cf. opposite page) meeting. I believe that research into tools and aids for human translators is a paradigm that has great potential which is far from reaching its limits. Parallel text alignment is one development that has had an impact, resulting in the new paradigm of translation memories. In general, people are looking for better adapted small technology aids that can be plugged into the whole multilingual production process. To develop such tools, you need to be willing to understand what those people's actual work process is about, and where the research you're involved in can fit in with a specific module, or a problem that needs to be solved.

For instance, it used to be the case that translators in large organisations used dictaphones to talk in a first version of a translation, which would subsequently be typed out by a secretary. For many translators, site translation into a dictaphone is the most practical and time-efficient way of approaching





the task. But when were inserted into the workflow, speech recognition was not an option. Translators were instructed to proceed directly to the written version, thus becoming their own secretaries in the process. But the latest developments in speech recognition suggest that we will soon be able to re-insert the speech part back into the document cycle, relieving the translator of the writing-up part. So we can now start to offer technological solutions, including speech-based ones while respecting the natural (and more efficient) human workflow. I see the development of work in speech as crucial to our whole field in providing a basis for more natural interaction with computers.

**ELSNNews:** Multilinguality is obviously an important issue in Switzerland. What is distinctive about research in this area in Switzerland?

**Armstrong:** For the EU community, multilinguality has become a popular buzzword. But in Switzerland, with its diverse linguistic communities, multilinguality has always been an important issue here, a way of life. Product information, for instance, always has to be given in at least two languages. In court, people have the right to defend themselves, and have access to information, in their own language. Avalanche bulletins must be available in all of the Swiss national languages, including English for the tourists. And so on. AT ISSCO, we strive to maintain linguistic competence in the four working languages in Switzerland — English, French, German and Italian. Multilinguality is a reality. But it does bring extra technological problems. I experienced this recently in our work with Swiss Telecom. Place names, for instance, are different in the different languages: this increases the problem of finding the correct match. In addition to this, people may be speaking in a language that isn't their native tongue, with an accent: that, too, can create confusion. From our perspective, developing adequate speech technologies for a single linguistic community, in countries like England and France, seems a lot simpler!

**ELSNNews:** We've had MULTEXT, ECI, ELRA: what are the important issues that still need to be resolved? Do we need more of the same work or something significantly new?

**Armstrong:** We need a lot more of this work, better quality and better distribution of the results!

One important need is for more data, and, particularly, focused collections of data in many languages that are limited by

domain, text type, etc. The most promising developments in technology for translation tasks rely on working in a limited domain. So what we need are comparable and parallel multilingual collections. For European evaluation actions we need similar data and also annotated collections in all languages. Now that ELRA is well in place, we can hope that their catalogue will continue to grow and provide these resources. Both LDC and ELRA provide an important service to our community, but there is also space for smaller spontaneous actions such as the ECI. Once the initial investment in resource development has been covered at a given site, these colleagues should be encouraged to share these resources with others.

We also need more work on basic core tools. What happens a lot at the moment is that different labs and companies develop their own tools - morphological analyzers, lexicons, taggers, statistical packages - often redoing what someone else has already done. We now have a baseline of core tools; instead of rebuilding them, we need to develop standard tools that can be shared.

My impression is that in Anglo-Saxon countries, there is a general philosophy that supports and even encourages making such things generally available. For English, a large number of lexical resources and tools are available. This is, unfortunately, not the case for other European languages. There may, in fact, be different national rules on what constitutes good research or academic behaviour.

Although EU funded projects aim towards (and have paid for) the development of shared tools and standards, the results have not always been what we've hoped for. This does not mean we should stop funding such work. And perhaps we need to take a long term perspective in assessing progress in this area, while appreciating the short term results that are available.

#### FOR INFORMATION

**Susan Armstrong** can be contacted at:  
 ISSCO, University of Geneva  
 54 route des Acacias  
 CH-1227 GENEVA (Switzerland)  
 Tel: +41 22 705 7113  
 Fax: +41 22 300 1086  
 E-mail: Susan.Armstrong@issco.unige.ch  
 WWW: <http://issco-www.unige.ch/>

LISA info

## The Localisation Industry Standards Association

LISA, The Localisation Industry Standards Association, is the world's premier organization for localization, internationalization and multilingual publishing. LISA organizes quarterly forums designed to facilitate an open exchange of ideas among the supply and demand sectors, and has a special interest group, OSCAR (Open Standards for Container/Content Allowing Re-use), which has just published its first results.

For more information on LISA's activities, please contact:

LISA administration  
 2 bis rue Ad-Fontanel  
 CH 1227 Carouge/Geneva  
 Switzerland  
 Tel: +41 22 301 5760  
 Email: [lisa@lisa.org](mailto:lisa@lisa.org)  
 WWW: <http://www.lisa.unige.ch>

Feb 1998

**elsnet**  
 .....  
 .....

# A Course in SpeechMania

*'A double room for two please, we'll be arriving next Friday'*

*Gerrit Bloothoof, Utrecht Institute of Linguistics OTS*

**SpeechMania** is a successful Philips software package for building spoken dialogue systems, well-known for its use by the Swiss and German railway companies for the development of their train travel information systems. The package includes components for speech recognition, speech understanding, dialogue management, lexicon management, language models, and speech output. It runs two applications simultaneously on a 200 MHz PC (MMX, 64 MB RAM) with a telephone interface board. That's all.

In January, six academics from the Netherlands and Northern Ireland attended a full week course in SpeechMania 2.1 at **Philips Aachen**. Our main interest in the course was to learn about the system in a way that would enable us to build spoken dialogue systems with our own students. We believe that the development of spoken dialogue systems will be a key element in future curricula in spoken language engineering; it is the area where speech technology and language technology meet, and this makes it of great educational value to students. But students don't just need to be able to develop systems: they also need to be aware of what the problems — solved and unsolved — are. We wanted to be prepared for questions from our students, and so we were keen to learn all about the background of SpeechMania. And, being lecturers ourselves, we inevitably kept a critical eye on the didactic approach of the course. In short, the event was a challenge for both participants and trainers. No doubt the trainers, **Mark Hrabak** and **Frank Sassenscheid**, who had ran ten previous courses (but only with customers from industry) had a difficult time with us, but they succeeded reasonably well in meeting our expectations.

The first day of the course gave an introduction to spoken language systems and the SpeechMania approach. For newcomers to Speech such an introduction would be indispensable; but for academics in speech technology it could have been presented in a more condensed way. The next two days were spent learning **HDDL**, the dialogue description language the course focuses on. Using the case study of a hotel room reservation system we learned, step by step, how to program actions, prompts, rules, implicit and explicit verifications and subdialogues, how to handle the status graph, and so on. We could check our own programmes in the off-line mode of SpeechMania, which was very helpful in debugging. But sometimes it was hard to keep pace with the trainer and identify the problems in our own software at the same time. We badly felt the need for a HDDL editor.

On the fourth day, we recorded our own prompts and built our recognition lexicon. The process of installing it took precious time, and we feel it could have been automated. The tools themselves were not always self-explanatory and were not backed up by on-line help (this will be solved in future versions of SpeechMania). We'll definitely need to study the manuals at home! But it worked, and at the end of the fourth day we had our system in place.

So after four days, there was the big moment when we could test our very first own application. It is absolutely exciting to see a system respond according to a dialogue flow of your own design. The dialogue design was tested beforehand in off-line mode, but it was revealing to see what happens when real speech recognition comes in. Our error handling was far from user-friendly and complete! The language model we trained on the few tests we did with each other's systems could not improve the systems' behaviour much. Still, we felt confident that with more investment in prototyping and training an acceptable system could be within reach.

The last day was spent mostly on the language model which was trained on some dialogues we had recorded and transcribed. And finally, we tested and evaluated the system again over the telephone (using US-English speaker-independent models and non-native speakers). Our initial systems had typical word error rates of 40% and concept error rates of 60%, and would need a lot of improvement!

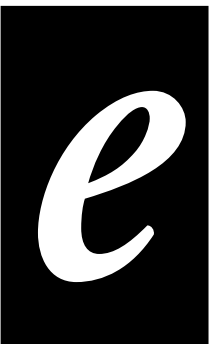
We were pretty tired afterwards. Looking back, we learned the basics of a new programming language (**HDDL**), and experienced lots of tools for the first time — something which can be confusing at times. The real proof of the pudding will come when we have to install the package at our universities and get all tools up and running. It is reassuring that **Philips Speech Processing** in Aachen won a Philips prize for best consumer service; I think we'll need it during the coming months.

To conclude, SpeechMania is a powerful package for fast initial development of spoken dialogue systems, but training and tuning of the system still takes a lot of time. Although SpeechMania has a modular structure, the recognition module cannot be replaced. In general one has to use the HMMs owned by Philips, which are available for various languages. The system prompts are pre-recorded, but could be replaced by a text-to-speech system of one's own choice. Academics can buy the system for 4000 DM, which includes attendance of the week-long course for two people — an absolute necessity.

## FOR INFORMATION

**Gerrit Bloothoof** can be contacted at:  
 Utrecht Institute of Linguistics OTS  
 Trans 10,  
 3512 JK Utrecht,  
 The Netherlands  
 Tel: +31 30 253 6042  
 Fax: +31 30 253 6000  
 Email: bloothoof@let.ruu.nl

For information on **SpeechMania**, please contact:  
 Christian Dugast  
 Philips Speech Processing  
 Kackertstrasse 10, D-52072  
 Aachen, Germany  
 Email: dugast@acn.be.philips.com



## Robustness: Real Life Applications in Language and Speech

13-24 July 1998  
Barcelona, Spain

*Robustness is perhaps the greatest single challenge for our understanding of speech perception and for speech technology. Speech-based courses at the Summer School will cover robustness issues in understanding informal speech, in robust automatic speech recognition and in the auditory system.*

*Robustness is also a challenge as far as natural language processing is concerned. NLP-based courses at the Summer School will deal with robust parsing techniques both for text and for spoken dialogues, and with topics which are highly relevant for commercial applications, such as style checking.*

### Structure and content

The Summer School will start every morning with a plenary session on the integration of Language and Speech. The topics of the plenary sessions are *Practical Natural Language Processing* during the first week and *The Processing of Spoken Language in Real World* during the second week.

Following the plenary session, there will be two slots of optional courses and practical workshops, run in parallel. The number of students in each parallel course will be limited, and course preferences must be indicated in the pre-registration form. The full Summer School programme is now available on the Web.

### Grants

A number of bursaries, provided by the TMR (Training and Mobility of Researchers) programme, are available. Terms and conditions can be found on the Web.

### Important dates

Pre-registration deadline:	March 15
Deadline for grant applications	April 1
Registration deadline:	May 1
Grant notification:	May 1
Payment deadline:	June 1

### FOR INFORMATION

Summer School Secretariat  
Dep. Teoria del Senyal i Comunicacions  
Universitat Politecnica de Catalunya  
Jordi Girona 1-3 MOD D5  
08034 Barcelona., Spain  
Fax: +34-3 401 64 47  
Email: summer98@gps.tsc.upc.es  
WWW: <http://gps-tsc.upc.es/veu/ess98/>  
or via ELSNET's Home Page: <http://www.elsnet.org>

ELSNET Summer  
School: Full  
Announcement

## Future events

Mar 16-20, 1998: *International Workshop on Intelligent Agents on the Internet and the Web*, Mexico City, Mexico. For more information, contact [s.murugesan@uws.edu.au](mailto:s.murugesan@uws.edu.au) URL: <http://btwebsh.macarthur.uws.edu.au/iaiw/>

Mar 23-25, 1998: *AAAI 1998 Spring Symposium on Intelligent Text Summarization*, Stanford, USA. For more information, contact [radev@cs.columbia.edu](mailto:radev@cs.columbia.edu) URL: <http://www.cs.columbia.edu/~radev/aaai-sss98-its>

Mar 25-27, 1998: *ELSNET in Wonderland*: conference, Soesterberg, near Utrecht, Netherlands. For more information, contact [elsnet@let.ruu.nl](mailto:elsnet@let.ruu.nl) URL: <http://www.elsnet.org/wonderland>

May 3-7, 1998: *ESCA Tutorial and Research Workshop on modeling pronunciation variation for automatic speech recognition*, Rolduc, Kerkrade, Netherlands. For more information, contact [STRIK@LET.KUN.NL](mailto:STRIK@LET.KUN.NL) URL: <http://lands.let.kun.nl/pron-var/>

May 13-15, 1998: *13th Twente Workshop on Language Technology*, Enschede, Netherlands. For more information, contact [joris@cs.utwente.nl](mailto:joris@cs.utwente.nl) URL: <http://www.cs.utwente.nl/Docs/parlevink/twlt/>

May 24, 1998: *Interaction Agents Workshop*, L'Aquila, Italy. For more information, contact [avi-ii@fub.it](mailto:avi-ii@fub.it) URL: <ftp://fub.it/pub/AVI-II98/>

May 25-27, 1998: *STiLL - ESCA-Workshop On Speech Technology In Language Learning*, Stockholm, Sweden. For more information, contact [still@speech.kth.se](mailto:still@speech.kth.se) URL: <http://ophale.icp.grenet/escal>

May 28-30, 1998: *First International Conference on Language Resources and Evaluation*, Granada, Spain. For more information, contact [lrec@ilc.pi.cnr.it](mailto:lrec@ilc.pi.cnr.it)

May 27, 1998: *Towards a European Evaluation Infrastructure for NL and Speech*. LREC Workshop, Granada, Spain. For more information, contact [steven.krauwer@let.ruu.nl](mailto:steven.krauwer@let.ruu.nl) URL: <http://www.icp.inpg.fr/ELRA/confle.html>

May 31-June 1, 1998: *The Japanese Localization Opportunity*. LISA Forum. Tokyo, Japan. For more information, contact [lisa@lisa.org](mailto:lisa@lisa.org) URL: <http://www.lisa.unige.ch>

Future Events

Feb 1998

**elsnet**  
.....

## ELSNET Secretariat

Steven Krauwer  
Coordinator

Mariken Broekhoven  
Assistant Coordinator  
Utrecht University (NL)

## Task Group

### Convenors

*Training & Mobility*  
Gerrit Bloothoof, Utrecht University (NL)

### Info Dissemination

Ewan Klein  
Edinburgh University (UK)

### Linguistic & Speech Resources

Antonio Zampolli  
Istituto di Linguistica Computazionale (I) and Ulrich Heid, Stuttgart University (D)

### Research

Niels Ole Bernsen  
Odense University and Joseph Mariani  
LIMSI-CNRS

## Industrial Panel

Harri Arnola, Kielikone (SF)  
Roberto Billi, CSELT (I)  
Michael Carey, Enigma (UK)  
Jean-Pierre Chanod, Rank Xerox Research Centre (F)  
Harald Höge, Siemens AG (D)  
Bernard Normier, GSI-ERLI (F)  
Brian Oakley (chair, UK)

## ELSNET Participants Academic Sites

NL Utrecht University (coordinator)  
A OFAI/Univ. Vienna/Vienna Univ. of Technology  
B University of Antwerp  
B University of Leuven  
BU Bulgarian Acad. of Sciences, Sofia  
BY Belarussian Academy of Sciences, Minsk  
CH IDSIA, Lugano  
CH ISSCO, Geneva  
CZ Charles University, Prague  
D Univ. des Saarlandes/DFKI, Saarbrücken  
D Univ. Hamburg  
D Univ. Kiel  
D Univ. of Stuttgart  
D Ruhr-Univ. Bochum  
D Univ. Erlangen  
DK Ctr for Sprogteknologie, Copenhagen  
DK Ctr for PersonKommunikation (CPK), Aalborg  
DK Odense University  
E Universidad de Granada  
E Univ. Politecnica de Catalonia/Univ. Autonoma de Barcelona  
E Univ. Politecnica de Madrid  
E Univ. Politecnica de Valencia  
F LIMSI-CNRS, Orsay  
F IRIT, Toulouse  
F Inst. de la Comm. Parlée, Grenoble  
F IRISA, Rennes  
F Laboratoire Parole et Langage-CNRS, Aix-en-Provence  
F CRIN, Nancy  
GR ILSP/NCSR "Demokritos", Athens  
GR Wire Communications Lab., Patras  
H Hungarian Acad. of Sciences, Budapest  
H Technical University, Budapest  
I Ist. di Linguistica Computazionale, Pisa  
I IRST, Trento  
I Fondazione Ugo Bordoni, Rome

IRL University College Dublin  
IRL University of Dublin  
IT Institute of Mathematics and Informatics, Vilnius  
N University of Trondheim  
NL Stichting Spraaktechnologie, Utrecht  
NL Inst. for Perception Research, Eindhoven  
NL Leyden Univ.  
NL Catholic Univ. of Nijmegen  
NL TNO Human Factors Research Institute  
NL Univ. of Amsterdam  
NL Univ. of Tilburg  
NL Univ. of Twente  
P INESC/ILTEC/Univ. Nova de Lisboa  
PL Polish Academy of Sciences, Warsaw  
RO Research Inst. for Informatics, Bucharest  
RU Russian Academy of Sciences, Moscow  
S KTH, Stockholm  
S Univ. of Linköping  
UK Defence Research Agency, Malvern  
UK UMIST, Univ. of Manchester  
UK Univ. of Cambridge  
UK Univ. College London/School of Oriental and African Studies (SOAS)  
UK University of Edinburgh  
UK Univ. of Essex  
UK Univ. of Dundee  
UK Univ. of Leeds  
UK Univ. of Sheffield  
UK Univ. of Sunderland  
UK Univ. of Sussex  
UK Univ. of Ulster  
UK Univ. of York  
D Novotech GmbH  
D pc-plus Computing  
D Philips Research Laboratories  
D Siemens AG  
D Verlag Moritz Diesterweg  
DK Tele Denmark  
E Telefonica I&D  
F ACSYS  
F Aerospaciale  
F GSI-ERLI  
F LINGA s.a.r.l.  
F MemoData  
F Rank Xerox Research Center  
F Systran SA  
F TGID  
F VECSYS Speech Processing  
GR Knowledge A.E.  
H Morphologic  
I CSELT  
I Database Informatica  
I Sogei (IRI-FINSIEL Group)  
I Syntax Sistemi Software  
I Tecnopolis CSATA Novus Ortus  
I Olivetti Ricerca SpA  
NL KPN Research Laboratories  
NL Polydoc N.V.  
NL University of Twente  
RU Analit, Ltd.  
RU Russicon Company  
S Telia Promotor (Call Centre Division)  
FIN Nokia Research Center  
FIN Kielikone Ltd  
UK ALPNET UK, Ltd  
UK BICC plc  
UK British Telecommunications  
UK Cambridge Algorithmica Ltd.  
UK Canon Research Centre Europe Ltd.  
UK Enigma Ltd.  
UK Hewlett-Packard Labs  
UK Logica Cambridge Ltd.  
UK Sharp Laboratories  
UK SRI International  
UK Vocalis Ltd.

## Industrial Sites

B Lernout & Hauspie Speech Products  
D aspect GmbH  
D Daimler-Benz AG  
D Electronic Publishing Partners GmbH  
D Grundig Professional Electronics GmbH  
D IBM Deutschland  
D Langenscheidt

## What is ELSNET?

ELSNET, the European Network in Language and Speech, was established in 1991 with funding from ESPRIT Basic Research. There were 25 founding members of the network. Currently, there are more than 60 universities and research institutes, and more than 45 companies participating.

The long-term technological goal which unites the members of ELSNET is to build integrated multilingual NL and speech systems with unrestricted coverage of both spoken and written language. Building multilingual NL and speech systems requires a massive joint effort by two pairs of communities: on the one hand, the natural language and speech communities, and on the other, academia and industry. Both pairs of communities are traditionally separated by wide gaps.

It is ELSNET's objective to provide a platform which bridges both gaps, and to ensure that all parties are provided with optimal conditions for fruitful collaboration. To achieve this, ELSNET has established an infrastructure for sharing knowledge, resources, problems, and solutions by offering (information) services and facilities, and by organising events which serve academia and industry in both the language and speech communities. In this respect, it is important to note that a network like ELSNET can only function well if all members of the network are prepared to give and to receive.

## Electronic Mailing List

elsnet-list is ELSNET's electronic mailing list. Email sent to [elsnet-list@let.ruu.nl](mailto:elsnet-list@let.ruu.nl) is received by all Managing, Associate and Industrial node coordinators of the Network, as well as other persons who have an interest in ELSNET's activities. This mailing list may be used to announce activities, post job openings, or discuss issues which are relevant to people in the European natural language and speech communities. To request additions/deletions/changes of address in the mailing list, send mail to [elsnet@let.ruu.nl](mailto:elsnet@let.ruu.nl).

## ELSNET web pages

Detailed information about ELSNET and its activities and publications is available on the Web at the following URL: <http://www.elsnet.org>. Comments and suggestions for new web pages are very welcome.

### FOR INFORMATION

ELSNET  
Utrecht Institute of Linguistics OTS, Utrecht University,  
Trans 10  
3512 JK Utrecht, The Netherlands  
Tel: +31 30 253 6039  
Fax: +31 30 253 6000  
Email: [elsnet@let.ruu.nl](mailto:elsnet@let.ruu.nl)  
WWW: <http://www.elsnet.org>

