elsnews

102

ELSNET Summer School goes to Prague

This year saw the 9th ELSNET Summer School on Language and Speech Communication held in Prague, and attended mostly by European students. This report has been written for ELSNews *by **Patricia Martin** of the Microsoft Corporation, who was one of five non-European participants.*

For this non-European attendee, the site of this year's programme on Text and Speech Corpora could not have been more perfect. Prague – the city of spires, churches, and history, with architecture ranging from gothic to romanesque, baroque to rococo, renaissance to art nouveau – was the ideal backdrop to this year's lectures. Nearly fifty students gathered from (largely) across Europe, with one or two from Korea, Israel, the USA, and South Africa, in Charles University – Prague's oldest university – to discuss the latest trends in our respective fields.

provided us with both Windows and Unix Labs to do our practical sessions. Each of the labs contained Czech language keyboards – for those who aren't used to it, the 41 letters of the Czech alphabet on the keyboards proved quite a challenge!

This sketch was presented to the organisers of the summer school as a gift from the participants

Charles University is situated in the *Malá Strana* section of Prague, or in English, the *Little Quarter*. The building itself shares a wall with the fantastic St Nicholas Church, one of the largest baroque churches in Europe, and the deservedly famous Charles Bridge, lined with ancient statues, was just a few hundred metres from the classroom.

The summer school was organised locally by the Institute of Formal and Applied Linguistics, and the Centre for Computational Linguistics, and was hosted by the Faculty of Mathematics and Physics, who

The courses covered the collection, validation, and use of a variety of types of data – text, speech, dialogue, and multimodal data. State-of-the-art techniques and tools were presented, and we had plenty of time for hands-on experience in the practical sessions.

The local host committee had done a wonderful job of organising the hotels, meals, and daily entertainment, with the latter ranging from brewery tours to historical tours of Prague castle. They were present throughout the day to answer questions, and worked night and day to make sure we enjoyed both the workshop and the city of Prague.

Summer 2001

elsnet

As might be expected, most of the students were from European universities, with a sprinkling of representatives from commerce and industry. The variety of native languages among the students attending were a linguist's dream, and a speech recogniser's challenge. English was the common language of the workshop, and all of us did our best to accommodate the various accents and pronunciations that were present. The students themselves were living examples of the prosodic variation present in a language. Teachers were from both sides of the Atlantic ocean, so this provided an excellent opportunity to learn about worldwide text and speech resources and practices.

The morning classes were followed by a two hour lunch break, which allowed students and teachers to mingle informally and continue the discussions raised during classes. Lunch was provided every day in a nearby restaurant, and most participants enjoyed the informal atmosphere and the chance to learn about each other's

*Some of the participants on Prague's famous Charles Bridge
Photograph by Maciej Karpinski*
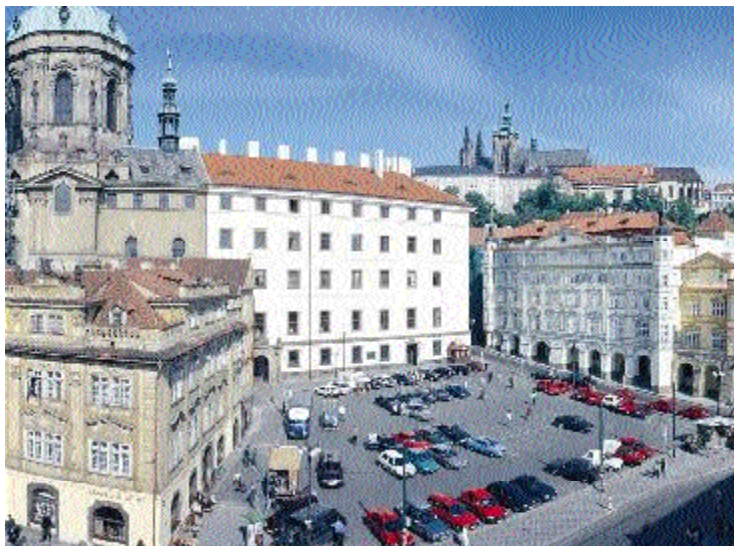
*The Malá Strana building of Charles University's Faculty of Maths and Physics – the venue of the summer school, showing the adjacent St Nicholas Church
Photograph from Charles University*

The second week's plenary session, taught by Uli Türk of the University of Munich, focused on the special requirements of multimodal corpora design principles and practices.

The future of speech recognition will include the use of additional information from the speaker – facial expressions, gestures, and other information. The second week's lecture, taught by three researchers from IBM, gave demonstrations and lab sessions to illustrate the problems of recognising facial information. Other classes and labs focused on annotations – at the prosodic, linguistic, and grammatical levels. Finally, after we had learnt how to create and annotate text, speech, and multimodal databases, we spent some time learning how to validate and verify the correctness of the data.

From corpus creation to annotation at any level; from sharing the data with others via useful metadata to learning how to evaluate the quality of a database; all of this knowledge was available at the ELSNET summer school. Get out your microphones, your video cameras, your word processors, and start producing!

work and research projects. Each of the afternoon sessions was followed by five student presentations: these were attended by both students and faculty, who were eager to answer questions and provide comments on the student topics.

This year's theme was Text and Language Corpora. The first week's plenary session, attended by all students, was taught by Steven Bird from the University of Pennsylvania. Steven's class informed us very clearly of what text and speech corpora are available for natural language processing, speech recognition, and text-to-speech research and development. The students attending this class left very well-informed about how to make their own resources available on the web through the use of metadata, and how to locate resources for their own work.

ReadIt! Applies a Revolutionary Approach to Reading Education

Luc Julia, *BravoBrava!*

bravobrava.com

BravoBrava! LLC has developed a piece of software to help children read. While the child reads aloud, the computer keeps the child on track and offers feedback when the child has difficulty. The feedback can be as subtle as changing the text colour for a well articulated phrase, or as friendly as a cartoon character that talks. The computer is infinitely patient and can keep detailed records of the child's progress. The reading software is being commercialised by BravoBrava's spin-off company, Soliloquy Learning, Inc.

soliloquylearning.com

As reviewed in the recent report of the American National Reading Panel, engaging children in supported oral reading is the most valuable method of working towards building their reading proficiency. At present, the only means of giving children such practice is by finding a human adult who will sit with them and help them. In order to break through this bottleneck, so that every child in every school can get the support that she or he needs, our vision is to use technology to provide a high-quality, low cost, automated reading coach that delivers voice-activated reading instruction, practice, and assessment over electronic media.
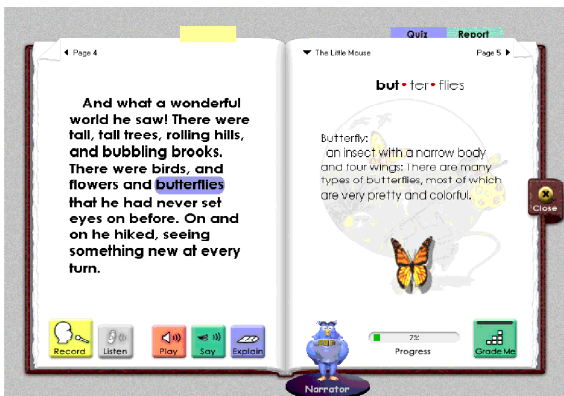
As the child reads, the system tracks the reading and intervenes when help is needed. This screen shows a simple intervention if the child stumbles on the word 'butterflies'

[*Editor's Note: The monochrome nature of this publication does not do justice to the original, which is very clearly colour-coded. For a demonstration of ReadIt! see BravoBrava's web site.*]

This technology will help reduce the digital divide, and provide an unprecedented level of tracking data to support teachers' instruction and assessment efforts, and to build the next generation of intervention techniques.

Here is a sample after the child has asked for a display of how well the selection was read. [Note previous comment on colour: in this example, 'red' words are 'what', 'bubbling', 'turn'; one black word, 'were'; yellow words – pale; all remaining are coloured green – Ed.]

The modular architecture allows for rapid reconfiguration according to specific needs. The current PC product uses Microsoft speech recognition software. However, the architecture supports the use of other recognisers, and we have experimented with others for use on different platforms. The audio outputs can be from recorded waveforms or, for maximum flexibility, can use a text-to-speech synthesis system. Any text can flow through the system for reading practice. Although explicit measures, such as a quiz to assess comprehension or vocabulary, can be included, several automatic measures are important by-products of use of the tool:

- Total number of words read by session, showing which were, and which were not, fluent
- Words per minute as a function of time
- Level of the material read as a function of time
- Number of times the intervention of the Reading Resource was used
- Number of times the child had the story read to him/her
- The actual recordings of what was read

Of course, such software will not help unless it gets into the hands of the children who need it. Therefore, an important strategy of the company is to support as many different platforms as possible to enable this goal to be met. The modular architecture of the system facilitates transition to various platforms. Similarly, to be of

Summer 2001

elsnet

maximal interest to the most children, we need as much appealing content as possible. In the area of human language technology, we hope such technology will evolve to automatically provide the following features:

- Context-specific definitions
- Context-specific and child-specific synonyms
- Rewriting of complex phrases into simpler ones

Solving these problems completely is a major research project. However, we believe that with careful matchmaking between the technical possibilities and the real needs of beginner readers, progress can be made in the near term.

Our mission is to dramatically improve learning by equipping teachers and parents with revolutionary technology products that permit significant improvements in education, are fun to use, and serve the broader social interest. We develop technologies and products that speed the delivery of established, effective reading techniques and offer a more efficient one-to-one and customised learning experience. We don't develop products that are a substitute for teaching or parenting. Our products are an ally of the teacher and the parent: we help teachers and parents extend themselves and their talents.

Success Story

ELSNET Supports Award Winner

Earlier this year ELSNET provided funding for Konstantinos Chandrinos from the Software and Knowledge Engineering Laboratory (SKEL) at the Institute of Informatics and Telecommunications at Demokritos – the National Centre for Scientific Research (NCSR) – in Athens, to attend the European Venture Academy seminar hosted by 'IZET, the Innovation Centre at Itzehoe, Germany, from 29-31 March.

*The product described by Konstantinos was **i-sieve** (see below), a spin-off activity of SKEL. The presentation won the **best business plan presentation** award of Euro 1,000, which went towards the registration for the follow-up Investment Forum, held in Milan from 26-28 April. The award was seen as recognition of the entrepreneurial potential of **i-sieve** for a European launch, and under pins the suggestion of the organising committee that the production team were ready to speak to investors and seek venture capital.*

*Konstantinos then applied for a further grant from ELSNET to enable him to attend the Investment Forum, where the SKEL plan **again** won the best business plan presentation award.*

ELSNews congratulates Konstantinos and the SKEL team, who are now working on a polished and informed business plan aimed at raising funding for i-sieve. We wish them luck.

The SKEL product, **i-sieve**, is an enabling technology that offers its users the opportunity to select content to match their needs and preferences. Major application fields for i-sieve include:

- *internet filtering*, where users may choose to fine-tune web queries or avoid unproductive, illegal, and harmful content for themselves or minors
- *e-mail filtering*, where i-sieve can be trained on mailboxes, to filter and/or categorise e-mail
- *e-news filtering*, where users can suggest examples of

can deduce such preferences from usage analysis and perform filtering with the users' approval
- *corporate filtering*, where companies may choose to prevent employees from accessing various external resources and/or distributing sensitive documents via electronic means, either accidentally or on purpose.

Systems incorporating i-sieve tools are trained offline on appropriate user-suggested and/or automatically-collected examples to identify multimedia features distinctive enough to infer the models for classes of documents.

i-sieve technology is adaptive to the language of the particular content, domain of application, or culture. The small footprint of the core technology makes it ideal for embedded solutions.

The overall impression of the SKEL team, and Konstantinos in particular, is that attending this type of venue is a very useful experience for future entrepreneurs, particularly if they do not have formal education in business and finance. In particular, those with technical or academic backgrounds benefit from being prepared for the different approach adopted by financiers when it comes to the valuation of a technical idea. It also provides a first-class opportunity to study how companies at various stages of financing are positioning themselves, and to have informal discussions about their problems.

ICAME, Where Will You Go?

Geoffrey Sampson, *University of Sussex*

Back in the 1970s, corpus linguistics was a minority activity, pursued by a tiny community who were seen by the vast majority of linguists and natural language processing experts as harmless eccentrics. The first computerised corpus of English had been produced at Brown University (Rhode Island, USA) by Nelson Francis and Francis Kučera in the 1960s, but it was widely ignored in its homeland; the influence of "generative", armchair approaches was so strong there that American linguists saw little point in working with quantities of real-life data. Corpora were a Northern European speciality: research happened mainly in Scandinavia, Britain, and the Netherlands, and excitement was centring round the British counterpart to the Brown Corpus, being compiled by Geoffrey Leech at Lancaster – what was to become the LOB (Lancaster-Oslo/Bergen) Corpus.

The completion of LOB was encountering difficulties, particularly to do with copyrights. Corpus enthusiasts saw a need to join forces in order to pool resources and move forward. One day in February 1977 Stig Johansson of Oslo convened a meeting attended by Nelson Francis (who was over on a sabbatical that year at Trondheim), Jan Svartvik of Lund, Jostein Hauge of Bergen, and Geoffrey Leech. This was the start of the first corpus linguistics society – ICAME, the *International Computer Archive for Modern English*. (When in due course ICAME developed a strong historical-linguistics group, the *M* was stretched to cover *Mediaeval* too.)

The participants enjoy a canal boat trip around Bruges
Photograph by Knut Hofland

Archive was in the name because a chief aim was to create a central point for collecting and distributing corpus resources – originally raw text on magnetic tape with microfiche concordances, nowadays texts with various levels of annotation on CD-ROM. This aspect

of ICAME's work has been led by Knut Hofland at the Norwegian Computing Centre for the Humanities, Bergen; for many years this was virtually the only source worldwide from which corpora were readily available, and ICAME continues to add new compilations to its catalogue – the latest offering, the COLT corpus of London teenager talk, comprises digitised speech signals as well as transcriptions. An Archive newsletter evolved into the *ICAME Journal*, and a lively and useful Corpora discussion list is hosted at Bergen (www.hit.uib.no/corpora/welcome.txt).

But the central activity of ICAME is surely its conferences. The first was in 1979, and since 1981 they have been held annually – usually in Europe, but recent years have seen meetings in Toronto and in Sydney. In May 2001 Sylviane Granger ran the 22nd ICAME Conference, with some welcome innovations in format, at Louvain-la-Neuve, Belgium.

Sylviane Granger and 'the Spanish contingent'
(Sylviane is 2nd from right on the top row)
Photograph by Sebastian Hoffmann

The Louvain-la-Neuve meeting brought to a head issues which had been raised last year at Sydney. In a sense, the original ICAME mission has been successfully accomplished. Linguistics departments may still contain a proportion of unreconstructed generativists, but no-one in charge of national or international research policies nowadays doubts the value of basing natural language research on electronic data samples. Several new organisations, notably the Pennsylvania-based Linguistic Data Consortium, as well as ELRA and the Oxford Text Archive, have become important resource-distributing centres. Corpus-based research is now an accepted feature of mainstream conferences, and new corpus-linguistics forums have emerged. LREC, focussing on the resources themselves, has been covered

*Some of the ICAME participants during an excursion to Bruges
Photograph by Knut Hofland*

date, ICAME has operated as the most informal of clubs. It has no constitution and no defined membership. Anyone is welcome to apply to come to a conference, but ICAME does not advertise, so the influx of newcomers has been slower than it could be. ICAME is steered by an Advisory Board (chaired for many years by Stig Johansson, and nowadays by Matti Rissanen of Helsinki) which consists roughly speaking of all people who have ever organised an ICAME conference (some older members no longer participate actively).

in previous numbers of *ELSNews*. A North American Association for Applied Corpus Linguistics held its third meeting this March. Later that month Tony McEnery of Lancaster organised the *Corpus Linguistics 2001* conference: this was devised initially as a one-off homage to Geoffrey Leech, the grandfather of European corpus linguistics, on the occasion of his retirement, but it proved so successful that it is likely to become a regular fixture, not restricted (as ICAME is) to English-language topics.
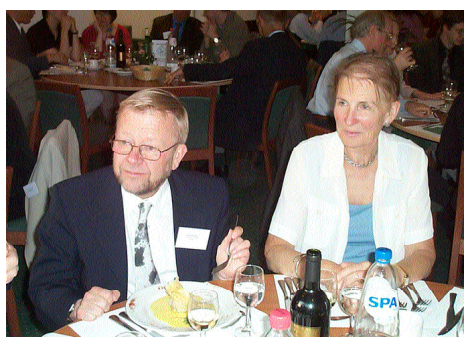
*ICAME President Matti Rissanen of Helsinki
with Anna-Brita Stenström from Bergen
Photograph by Knut Hofland*

Having started as the only game in town, ICAME now only partly covers the burgeoning field. If you want to publicise the availability of a new specialist corpus, or announce corpus-derived findings about the fate of some grammatical construction in the history of English or in foreign learners' struggles to master the language, ICAME is a good place to do it; but if your topic is performance of a corpus-based parsing algorithm, or the use of XML for corpus formatting standards, probably you would not choose ICAME. What should ICAME's role be in the 21st century?

Even more important at Louvain-la-Neuve than questions about purview were organisational issues. To

At Louvain-la-Neuve, some ICAMErs of long standing felt passionately that the "clubbable" atmosphere, in which most people at a conference know many or most others, was crucially valuable. It offers isolated researchers a sense of community and support which academics need as much as other professionals – this might be threatened by more competitive, American-style arrangements, in which meetings are promoted to maximise attendance, and the point of submitting a paper can be more to add another notch to one's tally of acceptances than to explore interesting work with a relaxed and like-minded audience. Others, though, felt equally strongly that ICAME will be by-passed as redundant, unless it adopts some degree of organisational reform, which will give it a mandate to identify and settle on distinctive ranges of activities in changing circumstances.

On the last morning at Louvain-la-Neuve, Matti Rissanen announced in outline what sounded like a fairly thorough acceptance by the Advisory Board of the "modernisers'" case, though implementation details will take some time to sort out.

Perhaps we should reflect that problems of this order are natural, when one generation's "eccentricity" comes to be recognised by the next generation as plain common sense.

FOR INFORMATION

Geoffrey Sampson is Professor of Natural Language Computing in the School of Cognitive and Computing Sciences at Sussex University, and is on the Editorial Board of *ELSNews*

Email: geoffs@cogs.sussex.ac.uk
Web: www.grsampson.net

For more information about **ICAME** visit www.hit.uib.no/icame.html

Letter to the Editor

ELSNews *has received the following letter, written under the pseudonym 'Swift Redux'. Readers interested in the relevance of this name may care to consider the title of the letter...*

To the Editor,

A Modest Proposal

As followers of the literature will have noted, great strides have been made in statistical parsing. In two decades, system performance figures have soared to over 90%. This is a magnificent tale. Parsing is cracked. An enormous debt is owed to the producers of the Penn Treebank. As anticipated by Don Walker, marked-up resources were what we needed. Once we had them, the algorithm boys could set to work, and whoomph!

The benefits of concentrating on the one corpus have been enormous. The field has focused. It has been the microscope under which the true nature of language has become apparent. Like Mendel unpacking the secrets of all species' genetics through assiduous attention to sweet peas, and sweet peas alone, Charniak, Collins, and others have unpacked the secrets of grammatical structure through rigorous attention to the *Wall Street Journal*.

We would now like to point out the unhelpfulness of comments appearing on the CORPORA mailing list, reporting low performance of various statistical POS-taggers when applied to text of different types from the training material, and also of a footnote to a recent ACL paper, according to which a leading Penn-Treebank-trained parser was applied to literary texts, but then its performance "significantly degraded". These results have not, I am glad to say, entered beyond that footnote into the scientific literature. The authors should realise that it is *prima facie* invalid to apply a resource trained on one type of data, to another. Anyone wishing to use a statistical parser on a text type for which a manually-parsed training corpus does not exist, must first create the training corpus. If they are not willing to do that, they may as well accept that ten years of dazzling progress is of no use to them.

We would like to contrast our exemplar of careful science with certain methodologies using the World Wide Web. An example: one experiment aims to find the English translation for French *groupe de travail*. A bilingual dictionary gives three translations for *groupe* and five for *travail*. Each of the combinations of one of the three and one of the five were sent to AltaVista, and its frequency on the web (according to AltaVista on that particular day) noted. 'Work group' had 67,238 hits, whereas the next most common had 1,131. This is presented as evidence that 'work group' is the correct translation.

We object. In scientific experiments, we expect some control over the inputs. How can it be scientific to present frequency contrasts of this order of magnitude, when the true scientists measure progress in fractions of percentage points? It is simply preposterous to base science on the web, wherein we know not what lies.

So now, our proposal. We are encouraged to see the amount of work based on the *Wall Street Journal* which appears in ACL proceedings. However, we remain concerned about the quantity of papers appearing there which fail to use a rigorous methodology, and fail to build on the progress outlined above. These papers tend to fall outside the domain which has become the testing ground for our understanding of the phenomenon of language, *viz*, the *Wall Street Journal*. Outside the *Wall Street Journal*, we are benighted. May I suggest that ACL adopt a policy of accepting only papers investigating the language of the *Wall Street Journal*.

Yours sincerely,
Swift Redux

Letters to *ELSNews*

At *ELSNews* we welcome letters from readers. Whether you wish to comment on anything appearing in this or other issues, give your opinion about matters relating to Human Language Technologies, or even if you want to ask an open question to other readers, please send your contributions to the Editor, Jenny Norris.

Email: jennyn@cogs.sussex.ac.uk

Please ensure that contributions are submitted by the deadline – for the next issue, this is 15 October.

Evaluation in Speech and Language Engineering: An Expert Bird's-Eye View

Patrick Paroubek, LIMSI – CNRS, Paris

Earlier this summer there were two events whose purpose was to discuss evaluation of speech and language technology systems. Patrick Paroubek was at both, and has given us these reports.

Patrick Paroubek

International Course on Speech and Language Engineering Evaluation

Despite quite a busy calendar at this time of the year, the 2nd and 3rd July saw this event held in Paris – jointly organised by CLASS (Collaboration in LAnguage and Speech Science and technology) and ELSNET.

During the course, a group of international experts from the field – some of them having been involved at the highest level for more than a decade in the largest evaluation programmes for speech and language technology worldwide – presented to a limited number of participants from both research and industry, the current state of deployment of evaluation in the field. Most of the people in the audience were more concerned with practical solutions than decision-making oriented issues.

Herman Steeneken's and Dave Pallett's presentations were punctuated by recounts of interesting anecdotes that brought a note of humour and were invaluable for conveying to the audience the hidden reality of large scale evaluation campaigns. John Garofolo (NIST) and Kathleen Stibler (Martin Lockheed) were, respectively, the advocates of the two classically opposed views on evaluations: namely, quantitative black-box technology oriented methods versus user-oriented ones. From Garofolo's presentation, it looks as if automatic meeting transcription will be a hot topic in the coming years.

Although coming from two different perspectives, both Phil Resnik's and Beith Sundheim's presentations addressed one of the essential issues of any evaluation in language processing – its grounding in human understanding/perception.

Throughout the course there were frequent interactions between the audience and the presenters: in particular, the issue of how to evaluate spoken language dialogue systems raised a lot of interest from the audience, since some of the attendees had come along specifically to find concrete answers to questions they are currently facing in their own work.

The general feeling was very positive. The audience expressed its gratitude to the organisers and the presenters for providing a comprehensive picture of the field, thus enabling those attending to get a quick grasp of the relationship between the various past, present, and future evaluation activities.

Workshop on Evaluation for Language and Dialogue Systems at EACL'01

Toulouse is reputedly sunny and very hot during the summer, but this time it was cool weather and rain that greeted the EACL'01 participants to the two parallel workshops that had evaluation on their agenda: the Senseval workshop, closing the current evaluation campaign on Word Sense Disambiguation; and the workshop on Evaluation for Language and Dialogue Systems, organised by David Novick (U. of Texas), Joseph Mariani (French Ministry of Research and LIMSI-CNRS), Candy Kamm (AT&T), Nils Dahlbäck (Linköping University), Frankie James (NASA), Karen Ward (U. of Texas), and myself. This two day workshop, held on 5th-6th July, with 38 registered participants, was split into three informal sessions: dialogue

systems evaluation; evaluation for language engineering in general; and a more focussed session on probabilistic issues and classification. These were followed by a final debate, amongst all the participants, on the current needs of the field in regard to evaluation.

The first sessions started with Anton Nijholt (U. Twente) giving a polished presentation, in which he offered us exciting views on multi-modal and multi-party contexts, illustrated with glimpses of avatars acting in their virtual environment. It raised the question brought forth by Dave Pallett (NIST), that is bound to come to the forefront of the scene in the coming years: how to perform evaluation in multi-modal communication? As Anton rightfully said, there is at present no answer to this question, leaving open a wide area of research since the problem is even more

complex than for spoken language dialogue systems. Tim Paek (Microsoft) could not be with us, so I took responsibility for presenting his paper entitled "Empirical Methods for Evaluating Dialogue Systems". In this paper Tim advocates the use of reference data built using a carefully crafted "Wizard of Oz" methodology as "Gold Standard" (this being the hypothesised maximal performance target for a given task), in conjunction with basic statistical metrics. The session ended with a presentation by Laila Dybkjær, in which she claimed that practical evaluation of walk-up-and-use spoken language dialogue systems should be performed by applying a set of evaluation criteria derived from the 15 usability issues which were identified as the result of the now-ended DISC project. Her presentation was followed by a short panel discussion with Frankie James, Anton Nijholt, Niels Ole Bernsen (NIS Labs), John Garofolo (NIST), and myself as panelists. Among the topics addressed during the discussion were: the dichotomy existing between user-oriented evaluation practices and black-box quantitative metrics; the fact that even without considering unrestricted dialogues, a dialogue whose scope reaches beyond simple booking tasks requires a good emulation of human understanding; and (mentioned by David Novick), the surprising fact that no paper presented during the workshop had raised the issue of a standard architecture for dialogue systems.

Anton Nijholt and Pierre Nugues using one of the breaks for some serious discussion
Photograph by David Novick

The second session grouped more diverse papers about evaluation. This started with a rather technical presentation by Stephen Watkinson (U of York) about the automatic translation of the Penn Treebank annotations into Categorial Grammar formalism, which raised the question of annotation standards for evaluation data, and of resource re-usability. It was followed by a presentation by Martine Hurault-Plantet (Limsi-CNRS) with a two level evaluation scheme applied to a system that participates in the Question & Answer track of TREC (Text REtrieval Conference). Then Widad Mustafa El Hadi (U. Lille 3) gave the audience an insight of the problems facing the organisers of a terminology extraction evaluation campaign, with her recounting of the ARC A3 evaluation campaign of the AUF (International association of French-speaking Universities).

The afternoon ended on a high note, with the duet presentation of Valerie Barr (Hofstra U.) and Judith Klavans (Colombia U.) who captivated the audience by

recounting the story of how they discovered that when linguists talk about evaluation, what they really mean is validation, in the language of software engineers.

In her invited talk for opening the last session, Donna Harman (NIST) advocated the benefits of focussed evaluation, using the examples of TREC and DUC (Document Understanding Conference). Then Yuval Krymolowski (Bar-Ilan U.) gave an interesting talk, where he showed that one can use the distribution of performance to study statistical NLP systems and corpora. Michel Jardino (Limsi-CNRS) concluded the session with her presentation of the comparison of two clustering methods (frequency-based collocation similarity association on the one hand and entropy-based on the other) of automatic topic identification or text classification. She claims that similarity of results can be used as validation criterion. A short panel discussion ensued between Donna Harman, Yuval Krymolowski, M. Jardino, Widad Mustafa El Hadi, and Martin Rajman (EPFL), with the problem of classifier evaluation in language engineering as an opening question.

The workshop ended with a general work session on the deployment of evaluation in language engineering. Joseph Mariani opened the discussion with a short picture of the current situation across the world (mostly the United States, Japan, and Europe). The general consensus was that standards and data are crucial assets for the development of evaluation in language engineering. In particular, the audience recognised the important role played by resource repositories like LDC and ELRA, as well as the need to have evaluation packages available. It was also said that an international framework should be installed in order to cooperate on promoting good practice in language technologies evaluation; on ensuring standard metrics, methods, and protocols; and on conducting studies in areas where the evaluation methods are still open (dialogue, spoken language translation...). The installation of a permanent entity in Europe, comparable to NIST, for the organisation of evaluation activities was recognised as an essential requirement of the field.

SIGdial
The USIX Interact Project: Adaptivity in Dialogue Systems

Kristiina Jokinen, University of Art and Design, Helsinki

One of SIGDial's goals is to promote research on discourse and dialogue, and to provide a forum for sharing results, tools, and components for exploring and building dialogue systems. This fourth SIGDial contribution to *ELSNews* presents **Interact**, a Finnish dialogue project aimed at developing models for natural and adaptive human-computer interaction.

The Interact project is a collaboration between four Finnish universities: University of Art and Design Helsinki (project coordinator); University of Helsinki; University of Tampere; and Helsinki University of Technology. It is funded by the National Technology Agency (TEKES), the leading IT companies (ICL Invia Oyj, Sonera Oyj, Lingsoft Oy, and Gurusoft Oy), as well as the Finnish Association of the Deaf (Kuurojen Liitto ry), and the Arla Institute (Arlainstituutti) – a vocational training and development centre for visually impaired and deafblind people. The three-year project started in June 2000 and will run until 2002.

The rapid development in IT has caused computer interaction to become more common, and systems should be able to adapt to various situations and users, so as to provide the most efficient and helpful mode of interaction. The aim of Interact is to explore natural human-computer interaction and to develop dialogue models which will allow users to interact with the computer in a natural and robust way. The research links up with (speech) user interfaces, and thus also concerns interaction management on several system levels.

The project deals with both speech and text communication, interfacing with the user by telephone, web-, and wap-applications. Various application types, such as intelligent question-answer systems, automatic call-centers, and other services that would benefit from flexible natural interaction will be investigated under the project.

Interact has so far built an initial integrated system with basic functionality and interaction capabilities, dealing with bus timetable inquiries. The current system accepts text input, whilst its output is spoken Finnish. The full demonstation system with spoken input/output modalities is scheduled to be released in spring 2002.

The innovative goal of this research is to facilitate natural language interaction in a wider range of situations than has so far been possible, and in situations where its use has not been sufficiently functional or robust. The need for flexible interaction is apparent not only in everyday computer use, but also in various situations where there is a need to automate services with the help of a dialogue system, and to make digital information accessible to all. This also implies that the special needs of disabled people will be taken into account when designing more natural interactive systems. Within the current system, such scenarios can, for example, include an intelligent bus-stop that allows speech and text interaction concerning city transportation.

The explosion of available information requires that systems deal with the problem of knowledge acquisition, and should be able to learn and adapt their knowledge to the changing situation. One of the important aspects of the project is to compare and test various soft-computing techniques aimed at automating and building a dialogue system that will also learn from its interaction with the user. In particular, unsupervised neural network techniques are used in the current system to determine what topic the user is talking about. The project also combines machine learning techniques with rule-based natural language processing, to investigate the limitations and advantages of the two approaches for language technology.

Interact also aims at stimulating discourse and dialogue related research more widely, and actively supports discussion of the development and design of practical dialogue systems capable of intelligent interaction.

FOR INFORMATION

Kristiina Jokinen is Project Manager of the Media Lab at the University of Art and Design in Helsinki

Email: Kristiina.Jokinen@uiah.fi

More information about the Interact project can be found at the home page, http://mlab.uiah.fi/interact/

For more information about SIGdial, visit http://www.sigdial.org

The Profession in 2001 – European Initiatives

John Nerbonne, *University of Groningen*

ELSNET links two research and development communities, each with its own professional society, and I've just returned from the ACL/EACL 2001, held in Toulouse, from 6-11 July. It was the 10th European Chapter Meeting and the 39th ACL Meeting. Of course, ACL is the Association for Computational Linguistics – the professional society for the language side of ELSNET's work. The Toulouse meeting was an impressive event, and here are some impressions.

The scientific papers were interesting both for their content and for indications they give of where the authors and referees now see the field. The trend toward a statistical approach to the classical questions of parsing and disambiguation continues, infused with new attention to linguistic detail. In this vein, there were papers on including sensitivity to grammatical heads in parsing, XML tools, maximising conditional versus joint probability, speech applications, and recovering dependency relations. There was new work on old applications such as machine translation, speech understanding, and question-answering, as well as papers on newer applications such as recognition and diagnosis of learners' errors, text summarisation, and information extraction on email. There were invited talks on linguistic search in audio files and on the use of statistical grammars in genetics. The prize for the best paper was shared between Eugene Charniak's paper on "Immediate-Head Parsing for Language Models" and a paper entitled "Fast Decoding and Optimal Decoding for Machine Translation" by Ulrich Germann, Michael Jahr, Kevin Knight, Daniel Marcu and Kenji Yamada. There were eleven workshops organised by ACL special interest groups (SIGs) on generation, dialogue, and learning, with groups covering topics ranging from data-driven machine translation, open-domain question-answering, and knowledge management to collocations, evaluation, and Arabic language processing. It was a high-energy place!

The meeting of the professional society is also interesting in terms of an overview of what's going on outside the conference. *Computational Linguistics*, the journal of the ACL, and the most important archive of the field's results, continues to attract so many submissions that only about 20% can be accepted. There are special interest groups for large corpora (SIGDAT), lexica, parsing, generation, learning (SIGNLL), dialogue, semantics, phonology, multimedia, and each of these has held special-purpose meetings in the last year or has published a CFP for a meeting soon. A new group devoted to Chinese Language Processing has just got started.

The ACL is embarking on a project to digitise all of the publications from its nearly forty years – not only the annual conference proceedings and annual and biannual chapter proceedings, but also all of the many SIG publications it distributes – 40k pages! COLING's publications are also distributed by the ACL, and COLING's organising committee has indicated that it will co-operate in this effort.

There were several indications that European organisational initiatives are breaking new ground and even setting trends. In an EACL panel on strategic funding for language technology, Joseph Mariani (French ministry and ELSNET Board Member) stood out for his clear vision separating fundamental research, technology development, and applications, while Bernd Reuse (German BMBF) disarmed scepticism about the (putatively missing) economic impact of language by calmly pointing to 20 products and eight companies that have spun off from Verbmobil. Nino Varile of *Directorate General XIII* of the European Commission foresees a place for end users and marketing closely integrated with research and development, and this seems to be unique among all the funding agencies. I was surprised that no-one mentioned a further point where the European Commission has been innovative in its strategy for language technology, and that is in creating a forum where academic and industrial R&D can exchange ideas and experiences in strategically inspired research – ELSNET. I know of no parallel at a national level.

The upcoming 40th anniversary of the ACL is also impressive. It might still be a bit arrogant to call computational linguistics a profession even *now*, putting it in a class alongside law, medicine, and civil engineering. Whoever had the chutzpah to call this a profession *39 years ago* must have been a daring soul indeed. But not crazy. In his Turing lecture, John Hopcroft recounted that, when he arrived in Stanford to teach computer science in the early sixties, no one could say very exactly what they wanted him to teach. Only years later did he recognise this as vision – a confidence in a strategic direction even before all the confirming details have crystallised.

FOR INFORMATION

John Nerbonne is Professor in the Department of Humanities and Computing at the University of Groningen in the Netherlands

Email: J.Nerbonne@let.rug.nl

**Web:** http://www.let.rug.nl/~nerbonne/

elsnet

HLT 2001 – Human Language Technology Conference

Jenny Norris, ELSNews *Editor*

March 18-21, San Diego, California

The setting could not have been more perfect – well, for the few of us who had left a flooded Britain behind, at least.

California via Arizona was dry and hot: the Catamaran Resort Hotel, the perfect setting, with its waterfall and exotic birds, not to mention its wonderful simultaneous balcony views over Mission Bay and the Pacific Ocean. The only trouble was that the Conference programme was so dense that our enjoyment of the
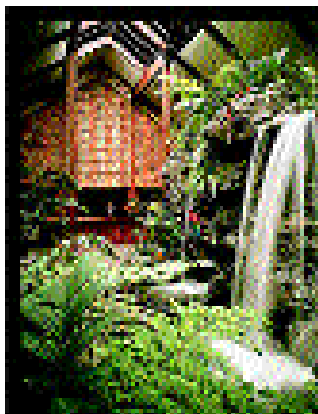
The hotel's main entrance lobby

setting was restricted to the gorgeous view from the balconies during coffee breaks. Conference sessions ran from 8.30 am, and we were occupied solidly until 9.30 pm, albeit with a banquet on one of the evenings.

As Editor of *ELSNews*, I had gone along to get an impression of the HLT scene in the USA, and to see if there are substantial differences between American and European approaches. Was it really true that American HLT concentrates largely on short-term, quick-fix applications, or was that just a vicious rumour I'd heard?

In general the programme was very well organised with sessions across the HLT board on: integrated systems; research issues (although this was really a chance for the American funders to tell the research community what they wanted to see); NLP; IE; dialogue systems; MT and cross-language systems; topic detection and tracking; and summarisation and question answering.

A distinction that did seem clear was the greater emphasis on speech applications across the USA – which may be a result of the priorities of the funding agencies. Certainly, the development of **generic** speech recognition technology is notoriously difficult – with enormous manual requirements and problems in porting to different tasks or languages.

These difficulties were the subject of a paper presented by Lori Lamel on behalf of the Spoken Language Processing Group at LIMSI-CNRS in Paris, who discussed possible methods of reducing the manual effort required in system development, and assessing the genericness of wide-domain models.

So what of the "superficiality" rumour I'd heard. Faced with the problems of portability and genericness, it is easy to see why the tendency has been to develop application-specific software. It does seem that without some fairly significant breakthroughs in technology (and methodology), the applications being developed – particularly in relation to speech recognition – will be limited. And that does beg the question of whether it is really worth spending such enormous sums on individual quick-fix applications.

As one of the few Europeans there (25 of the 260 participants represented Europe, with a similar number from Asia), I found the constant involvement of, and repeated reference to, the funding agencies rather irritating (the event was partly funded by DARPA and NSF), with their representatives clearly wielding considerable power and influence across the whole field. It was my impression that the Americans were not unduly bothered by this aspect, although I suppose they are just used to it.

On the whole, there were some very interesting talks. In his keynote speech, Fernando Pereira was pessimistic about the validity of using structured documents as normal data for corpus work: they have many of the same types of ambiguity as natural language, and add an extra layer of ambiguity of their own. He was therefore advocating the use of web language as natural language – a popular thesis at present, but one about which objections are being raised. (And see page 7 of this issue!)

The poster and demonstration sessions were well organised (apart from presentation boards being smaller than people had been told in advance), with "boaster sessions" preceding them, in which presenters had one microphone-enforced minute to convince delegates to visit their slots. This seemed to work very well indeed, and gave a good overview from which we could select what we wanted to see.

As for my other reason for being there, I found that a large number of participants had never come across *ELSNews* before: indeed, many had not heard of ELSNET until Joseph Mariani's talk, which gave a summary of European programmes and national initiatives in HLT. There was a lot of positive reaction to the principle of international networks of excellence – and I was happy to see that the number of American subscriptions to *ELSNews* went up after the event.

FOR INFORMATION

To see the full HLT 2001 programme, and to see how full the itinerary was, visit www.hlt2001.org/

MorphoLogic's First Decade

Szabolcs Kincse, Morphologic

In 1991 in Hungary the heavy restrictions on information and technology transfer had just been lifted by the West, and the country was facing a boom in the IT industry. The demand for the long-proposed Hungarian spell-checker was increasing, but official research brought only promises.

Hungarian, a Finno-Ugric language – like Finnish or Estonian – is an extremely productive language. This makes normal spell-checking processes virtually impossible due to the immense size of the necessary database. One average dictionary with a hundred thousand stems could produce billions of grammatically correct word forms – which is something that traditional spell-checking processes simply could not hope to deal with.

The solution to this problem lay in morphological analysis, and the then newly-formed MorphoLogic developed a morphology-based spell-checker and began licensing their product to many major actors in the market – such as Microsoft, Lotus, Xerox, Franklin, etc.

This year the company is ten years old and now provides linguistic search-enhancement solutions for Hungarian and other Eastern and Western languages. The necessity of incorporating stemming into the traditional search process is now widely recognised, and MorphoLogic's language-independent approach has enabled them to support many languages at the same time, so that they now provide linguistic support for many search engines.

The company has recently developed its no-click bilingual dictionary software, that determines the language of the word pointed at on the screen, and translates it into the required language. Even the conversion of kanji (a Japanese system of writing in Chinese characters) is fast. The method is based on morphological and local syntactic analysis, which allows recognition not just of the base forms of words, but also of split compounds and expressions. This software won a European Information Technology (IST) Prize in 1999, and today many dictionaries have been published using this format – under the name of **MoBiMouse**, or, most recently, as **Pointer**. MoBiMouse dictionaries are now the most popular types of e-dictionaries in Hungary.

State funding is now available for the research of human language technologies in Hungary, due to the extensive support (the so-called Széchenyi Plan) of the Hungarian government. Consequently, MorphoLogic is now running a joint project with the Research Institute of

Linguistics of the Hungarian Academy of Sciences, Szeged University, and the Gallup Institute – and they are jointly developing a method to process political and economic news items using content-based search that is based on XML technology. For the purposes of analysing the content of news, they plan to develop a parser capable of recognising complex patterns in the source text. Using lexical constraints in the patterns can facilitate the description of idiomatic expressions. An important advantage of the proposed system is that the set of possible attributes is not pre-determined. The formalism for describing a grammar is expected to become available during the summer of 2001.

MorphoLogic has also been awarded research funding for the morphological description of a number of endangered Finno-Ugric languages. This is long overdue – these peoples live on the brink of Europe, mostly around the Urals, and, due to the extensive assimilation activities, some of them are completely disappearing. MorphoLogic has joined forces with the Research Institute of Linguistics of the Hungarian Academy of Science, and together we are using the company's technology to document these distinctive languages, some of which are closely related to Hungarian or Finnish.

Other current research projects at MorphoLogic include: language-independent machine translation; creating linguistic support for OCRs, speech and writing tools, POS disambiguation, and knowledge management. We are developing our own integrated concordancer, to be completed by late 2002/early 2003. This integrates many NLP functions, from part-of-speech tagging to stemming (which facilitates the collection of statistics for languages with complex morphology), for many languages. And last, but not least, we are building up a complete NLP toolkit (comprising mostly in-house tools, with some from external sources) to help in the development and compilation of linguistic databases – from dictionaries, through morphology lexicons, to grammars for parsing.

FOR INFORMATION

Szabolcs Kincse is International Relations Manager at Morphologic

Email: kincse@morphologic.hu
Web: www.morphologic.hu

Summer 2001

elsnet

SALTMIL ISCA SIG
Speech And Language Technology for Minority Languages

Kepa Sarasola, *University of the Basque Country*

In ELSNews 10.1, which was dedicated to work on minority languages, there was not sufficient space to include a report on the activites of the SALTMIL Special Interest Group of ISCA. We promised to include a summary of the group's work in this issue, and here it is ...

Aims

The ISCA (International Speech Communication Association) Special Interest Group on Speech and Language Technology for Minority Languages has the overall aim of promoting research and development in the field of speech and language technology for lesser-used languages.

Activities

The SALTMIL web site (http://isl.ntftex.uni-lj.si/ SALTMIL) shows the group's aims, activities, history, membership information, literature references, and links to other similar resources.

The group runs an email list, accessible at http://www.egroups.com/group/saltmil. Messages are mostly from members, stating which minority language they are working on, and discussing their work and interests.

In May 1998, a workshop was held in Granada, Spain, on the theme of "Language Resources for European Minority Languages". A report with papers for downloading and a review of the workshop written by Nicholas Ostler are accessible from our web site.

In May 2000 a workshop preceding LREC 2000 was held in Athens, Greece. Its title was "Developing language resources for minority languages: re-useability and strategic priorities" (http://isl.ntftex.uni-lj.si/ SALTMIL/lrec00.html). The workshop proceedings are available from ELRA, ITE (the Linguistic Institute of Ireland), and our web site.

In November 2000, SALTMIL participated in the 2nd International Multimedia and Minority Languages Congress (http://www.gaia.es/multilinguae), held in the Miramar Palace in Donostia-San Sebastian.

Speech And Language Technology for MInority Languages

Future Activities

SALTMIL is one of the SIGs presenting in the SIGshow of Eurospeech 2001 Scandinavia, to be held in Aalborg (Denmark) in October.

3rd SALTMIL workshop to be run simultaneously with LREC 2002 in Las Palmas, Canary Islands, Spain, on 27 or 28 May, 2002.

SALTMIL will be participating in the 2002 Summer School on Speech and Language Processing for Local Languages.

SALTMIL will be creating a journal with two main features: merging speech and NLP areas; and giving special emphasis to issues concerning less prevalent languages .

FOR INFORMATION

Kepa Sarasola is a member of the IXA Group for NLP in the Department of Computer Languages and Systems at the Unversity of the Basque Country, and a member of the SALTMIL group

Email: jipsagak@si.ehu.es
SALTMIL web site: http://isl.ntftex.uni-lj.si/ SALTMIL

Future Events

Sept 3-7   *7th European Conference on Speech Communication and Technology (Eurospeech 2001 – Scandinavia)*: Aalborg, Denmark. Email: gosta.bruce@ling.lu.se; URL: http://eurospeech 2001.org

Sept 5-7   *Recent Advances in Natural Language Processing (RANLP-2001)*: Tzigov Chark, Bulgaria. Email: galia@lml.bas.bg;   URL: http://lml.bas.bg/ranlp2001/

Sept 7-9   *3rd Biennial Conference on Practical Applications in Linguistic Corpora (PALC 2001)*: Lódz, Poland. Email: corpora@krysia.uni.lodz.pl;   URL: http://www.uni.lodz.pl/pelcra

Sept 9-13   *24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001)*: New Orleans, U.S.A. Email: ddlewis2@worldnet.att.net;   URL: http://www.sigir2001.org

Sept 13   *Workshop on Operational Text Classification Systems 2001* (with SIGIR 2001): New Orleans, U.S.A. Email: otc2001info@DavidDLewis.com; URL: http://www.DavidDLewis.com/events/otc2001

Sept 14-15   *2nd International Workshop on Spanish Language Processing and Language Technologies (SLPLT-2)*: Jaén, Spain. Email: david@crl.nmsu.edu;   URL: http://sepln2001.ujaen.es

Sept 18-22   *Workshop on Teaching Machine Translation* (with MT Summit VIII): Santiago de Compostela, Spain. Email: tmt@dlsi.ua.es;   URL: http://www.dlsi.ua.es/tmt/

Sept 18-22   *8th Machine Translation Summit (MT Summit VIII)*: Santiago de Compostela, Spain. Email: summitVIII@eamt.org;   URL: http://www.eamt.org/summitVIII

Sept 20-22   *7th annual conference on Architectures and Mechanisms for Language Processing (AMLaP-2001)*: Saarbruecken, Germany. Email: amlap2001@amlap.org; URL: http://www.amlap.org/2001/

Sept 20-23   *5th FEL Conference – Endangered Languages and the Media*: Agadir, Morocco. Email: hasouzz@casanet.net.ma;   URL: http://www.ogmios.org/felmedia.htm

Sept 23-28   *4th International Tbilisi Symposium on Language, Logic and Computation*: Borjomi, Georgia. Email: chiko@contsys.acnet.ge;   URL: http://www.illcuva.nl/borjomi

Sept 24-26   *4th International Conference on The Electronic Document (CIDE 2001)*: Toulouse, France. Email: cide2001@irit.fr;   URL: http://www.irit.fr/CIDE2001

Oct 17-19   *7th International Workshop on Parsing Technologies (IWPT'01)*: Beijing, China. Email: wanghf@pku.edu.cn;   URL: http://www.icl.pku.edu.cn/iwpt2001.html

Oct 22-24   *ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding* Red Bank, NJ, U.S.A. Email: julia@research.att.com;   URL: http://ssli.ee.washington.edu/conferences/prosody

Nov 30   *12th Meeting of Computational Linguistics in the Netherlands (CLIN 2001)*: Twente, The Netherlands. Email: clin@cs.utwente.nl   URL: http://parlevink.cs.utwente.nl/Conferences/clin2001.html

Dec 3-5   *Information Society Technologies (IST 2001) – Technologies Serving People* (European Commission Event): Düsseldorf, Germany.   Email: infso-ist2001@cec.eu.int
URL: http://www.europa.eu.int/information_society/newsroom/istevent/programme/index_en.htm

Dec 9-13   IEEE 2001 *Automatic Speech Recognition and Understanding Workshop (ASRU'01)*: Madonna di Campiglio (Trento), Italy. Email: asru01@itc.it; URL: http://asru01.itc.it

Dec 11-13   *IRCS Workshop on Linguistic Databases*: Philadelphia, USA. Email: sb@ldc.upenn.edu   URL: http://www.ldc.upenn.edu/annotation/database/

Summer 2001

elsnet

This is only a selection of events – see http://www.elsnet.org/cgi-bin/elsnet/events.pl for details of more events

ELSNET

Office
Steven Krauwer,
Co-ordinator
Brigitte Burger,
Assistant Co-ordinator
Monique Hanrath,
Secretary
Utrecht University (NL)

Task Groups
Training & Mobility
Gerrit Bloothooft,
Utrecht University (NL)
Koenraad de Smedt,
University of Bergen (NO)

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

Research
Niels Ole Bernsen, NIS
Odense University (DK)
and Joseph Mariani,
LIMSI-CNRS (F)

Executive Board
Steven Krauwer,
Utrecht University (NL)
Niels Ole Bernsen, NIS,
Odense University (DK)
Björn Granström,
Royal Institute of Technology (S)
Nikos Fakotakis,
University of Patras (EL)
Ulrich Heid,
Stuttgart University (D)
Joseph Mariani,
LIMSI/CNRS (F)
José M. Pardo,
Polytechnic University of Madrid (E)
Geoffrey Sampson,
University of Sussex (UK)
Antonio Zampolli,
University of Pisa (I)

ELSNET Participants

Academic Sites

A	Austrian Research Institute for Artificial Intelligence (ÖFAI)
A	Graz University of Technology
A	University of Vienna
A	Vienna University of Technology
B	Leuven University
B	University of Antwerp - UIA
BG	Academy of Sciences Institute of Mathematics
BY	Belorussian Academy of Sciences
CH	SUPSI University of Applied Sciences
CH	University of Geneva
CZ	Charles University
D	Christian-Albrechts University, Kiel
D	German Research Center for Artificial Intelligence (DFKI)
D	Institute of Applied Information Science (IAI)
D	Ruhr-Universität Bochum
D	Universität Erlangen Nürnberg - FORWISS
D	Universität Hamburg
D	Universität Stuttgart
D	Universität des Saarlandes
DK	Aalborg University
DK	Center for Sprogteknologi
DK	University of Southern Denmark
E	Polytechnic University of Catalonia
E	Universidad Nacional de Educación a Distancia (UNED)E
E	Polytechnic University of Madrid
E	Polytechnic University of Valencia
E	Universitat Autonoma de Barcelona
E	University of Granada
EL	Institute for Language & Speech Processing (ILSP), Athens
EL	NCSR 'Demokritos', Athens
EL	University of Patras
F	IRISA/ENSSAT, Lannion
F	Inst. National Polytechnique de Grenoble
F	Institute de Phonétique, CNRS
F	LIMSI/CNRS, Orsay
F	LORIA, Nancy
F	Université Paul Sabatier (Toulouse III)
GE	Tbilisi State University, Centre on Language, Logic and Speech
HU	Lóránd Eötvös University
HU	Technical University of Budapest
I	Consiglio Nazionale delle Ricerche
I	Consorzio Pisa Ricerche
I	Fondazione Ugo Bordoni
I	IRST, Trento
I	Università degli Studi di Pisa
IRL	Trinity College, University of Dublin
IRL	University College Dublin
LT	Institute of Mathematics & Informatics
NL	Eindhoven University of Technology
NL	Foundation for Speech Technology
NL	Leiden University
NL	TNO Human Factors Research Institute
NL	Tilburg University
NL	University of Amsterdam
NL	University of Groningen
NL	University of Nijmegen
NL	University of Twente
NL	Utrecht University
NO	Norwegian University of Science and Technology
NO	University of Bergen
P	University of Lisbon
P	INESC, Lisbon
P	New University of Lisbon
PL	Polish Academy of Sciences
RO	Romanian Academy
RU	Russian Academy of Sciences, Moscow
S	KTH (Royal Institute of Technology)
S	Linköping University
UA	IRTC UNESCO/IIP
UK	Leeds University
UK	SOAS, School of Oriental and African Studies
UK	UMIST, Manchester
UK	University College London
UK	University of Brighton
UK	University of Cambridge
UK	University of Dundee
UK	University of Edinburgh
UK	University of Essex
UK	University of Sheffield
UK	University of Sunderland
UK	University of Sussex
UK	University of Ulster
UK	University of York

Industrial Sites

B	Lernout & Hauspie Speech Products
D	Acolada Gmbh
D	DaimlerChrysler AG
D	Grundig Professional Electronics GmbH
D	IBM Deutschland
D	Langenscheidt KG
D	Novotech GmbH
D	Philips Research Laboratories
D	Sympalog Speech Technologies AG
D	Varetis Communications
D	aspect Ges. für Mensch-Maschine Kommunikation mbH
DK	Tele Danmark
E	Sema Group sae
E	Telefonica I & D
EL	KNOWLEGDE SA
F	Aerospatiale
F	LINGA s.a.r.l.
F	LexiQuest
F	Memodata
F	SCIPER
F	Systran SA
F	TGID
F	VECSYS
F	Xerox Research Centre Europe
FIN	Kielikone Oy
FIN	Nokia Research Center
HU	MorphoLogic Ltd
I	LOQUENDO
I	OLIVETTI RICERCA SCpA
LV	TILDE
NL	Cap Gemini Nederland BV
NL	Compuleer
NL	IP Globalnet Nederland BV
NL	KPN Research
NL	Knowledge Concepts BV
NL	Sopheon
RU	ANALIT Ltd
RU	Russicon Company
S	Sema Group Infodata
S	Telia Promotor
UK	20/20 Speech Ltd
UK	ALPNET UK Limited
UK	BT Adastral Park
UK	Canon Research Centre Europe Ltd
UK	Hewlett-Packard Laboratories
UK	Imagination Technologies plc
UK	Logica Cambridge Ltd
UK	Reuters Ltd
UK	SRI International
UK	Sharp Laboratories of Europe Ltd
UK	Vocalis Ltd

What is ELSNET?

ELSNET, the European Network of Excellence in Human Language Technologies, is funded by the European Commission's Human Language Technologies programme. Members are academic and public research institutes (81) and industrial companies (55) from all over Europe.

The long-term technological goal, which unites the members of ELSNET, is to build integrated multilingual natural language and speech systems with unrestricted coverage of both spoken and written language. However, the realistic prospect for commercial applications involves systems that are restricted in one way or another. Such systems are of crucial importance for Europe in that they allow implementation of, and access to, the emerging multilingual information infrastructure. These systems also contribute to the increase of European industry's competitiveness by giving better access to product and service markets across language barriers.

Building multilingual language 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

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 the language and speech communities.

Electronic Mailing List

elsnet-list is ELSNET's electronic mailing list. Email sent to elsnet-list@let.uu.nl is received by all member site contact persons, as well as other interested parties. This mailing list may be used to announce activities, post job openings, or discuss issues which are relevant to ELSNET. To request additions/deletions/changes of address in the mailing list, please send mail to elsnet@let.uu.nl

Subscriptions

To subscribe to *ELSNews* visit **http://www.elsnet.org** and follow the links to *ELSNews* and subscription.

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@elsnet.org
Web: http://www.elsnet.org