

## New Swedish Graduate School for Language Technology

*Anders Green, Royal Institute of Technology, Stockholm*

*"A high quality graduate education that will become an important factor for research in natural language processing."*

This is the bold goal of GSLT, the recently opened Swedish National Graduate School in Language Technology.

The grand opening of GSLT was held on September 13 in Göteborg on the west coast of Sweden. About one hundred people attended the opening ceremony, which took place in the "Humanisten" building in Göteborg. The opening speakers were the Dean of the Faculty of Arts, Lennart Olausson, and the Secretary of State for the Swedish Education Department, Agneta Bladh. They expressed the view that, although the Swedish language is small in comparison to the big European languages such as English, French, and German, initiatives like the GSLT will make it possible to obtain language resources like spell-checkers, grammar-checkers, and knowledge management for Swedish as well.

GSLT is one of 16 graduate schools that have been funded by the Swedish government, with funding for seven years. It is a collaboration between leading centres in language technology and some regional universities, and is hosted by the Faculty of Arts at Göteborg University. The director is Robin Cooper, with Joakim Nivre (Växjö University) and Lars Ahrenberg (Linköping University) as deputy directors. At the opening

ceremony Robin Cooper stressed the importance of having a graduate school of high international standard dedicated to natural language technology.

The opening of GSLT began with two invited talks: Professor Hans Uszkoreit (Universität des Saarlandes) gave a lecture entitled "Language Technology for the Knowledge Society", in which he pointed out the importance of natural language processing within a knowledge-based society. Dr Lori Lamel (LIMSI-CNRS) talked about "Spoken Language Processing for Human-Machine Communication". Speech processing

has now matured to a point where it can be used for some interesting applications, and one of Dr Lamel's demonstrations showed the use of speech processing techniques for indexing audio broadcasts. She also showed a video presenting speech interfaces for public information kiosks used for giving public transport information.

Talks like these are a part of the graduate school's long-term goal of providing interesting topics for seminars in conjunction with graduate courses.

### Twenty five graduates in seven years

The administrative status of the graduate school is not uncomplicated. The Swedish education system requires that graduate students be admitted to a university that has the right to grant doctoral degrees. Since GSLT is not a university, it has no formal right to do this.



*From left to right, professors Lars Ahrenberg and Robin Cooper, deep in discussion with Joakim Nivre and Torbjörn Lager*

**SYNFACE:**  
a Communication Aid  
for Hearing-Impaired  
People  
Andrew Faulkner 3

**RACAI**  
Dan Tufis 4

**MT Summit VIII**  
Andrei Popescu-Belis 6

**ELSNets Interviews**  
Roger Moore of  
20/20 Speech Ltd 7

**SiGdial**  
Hans Dybkjær 9

**BalkaNet**  
Kemal Ofizer  
Dimitris Christodoulakis  
Sofia Stamou 10

**Lexicom Masterclass**  
Sue Atkins 11

**Opinion Column**  
John Nerbonne 13

**Imagination 2001**  
Gerrit Bloothoof 14

**Editor's Note** 14

**Call for Letters** 14

**Future Events** 15

**Autumn  
2001**

**elsnet**  
.....

Instead, the graduate school provides the funding for the graduate students, who follow a programme at their home university that conforms to GSLT requirements. Admission will be once a year for the first three years, with thirteen students admitted this year, and the goal of 25 students graduating over the next seven years.

At the moment the participating universities are: University College of Borås, University of Skövde, Växjö University, Chalmers University of Technology, Royal Institute of Technology, Linköping University, and Uppsala University. Negotiations are under way with other universities, and also with representatives from the industrial sector, such as SICS, the Swedish Institute of Computer Science.

### Workshop and retreat

After the opening ceremony all the supervisors and graduate students went to Lökeberg, outside Göteborg, for a three-day retreat and workshop devoted to presenting research at the participating departments. The students presented their research topics, and discussed their expectations of the school. Even though the spectrum of research interests was wide, some main trends could be spotted, ranging from speech processing, multimodal interaction, and information retrieval, to machine translation.

Most people at the workshop shared the view that the graduate school will be an important factor in the study of language technology in Sweden. By being a big virtual department of language technology it will be possible to achieve more than can be done by an individual department.

One topic of discussion was how students could benefit from the different disciplines that the graduate school comprises. The multidisciplinary nature of natural language processing means that everyone involved acknowledges that other disciplines have a different way of looking at things, perhaps using the same terms for the same phenomena.

### Collaboration within a virtual organisation

Another topic of discussion was how GSLT, being a "strange administrative creature", could work as a virtual organisation. It was clear that some kind of web-based system for collaborative work should be used, and since the computer resources of the school are quite good (and all students have their own laptops), this should not prove too problematic. It is still early days and students and staff (including the administrative staff) are all getting used to being part of the new school. But as the GSLT web site and the home pages of the graduate students develop, we will hopefully see the development of a useful tool for discussion of research questions, and a valuable source of information for language technology.

The strength of having representatives of different universities is that members of the graduate school will be able to share information about events like seminars, workshops, and research projects that otherwise could prove difficult to obtain. With this in mind, one of the decisions made at the workshop was that there should be a person responsible for providing the GSLT and its participants with information.

### A broader spectrum of courses

Immediately after the retreat the graduate school kicked off with one week of intensive studies: two courses were taught, aimed at providing students with an up-to-date picture of the state-of-the-art within natural language processing and speech processing.

One of the important features of GSLT is that its relatively large size makes it possible to give a broader spectrum of courses than would otherwise be possible. An important result of this is that it is now possible to plan courses that have other courses as prerequisites – something that will facilitate depth of study within a topic as well as having full lectures.

The courses given by GSLT are also available to graduate students who are not at the school. Consequently, about ten students from the Nordic countries were able to participate in the two first courses. These students received travel support from NorFa, a Nordic institution whose goal is that the Nordic countries be a common region for research training.

The practical work within the school will be concentrated into one or two weeks of intensive "residential" lectures, with students staying at Göteborg, after which the students will continue course work at their affiliated institutions.

At the end of the courses, students will gather to present their course work at the institution of the home department. At these intensive gatherings there will be seminars or lectures, with invited speakers, as with Hans Uszkoreit and Lori Lamel on this occasion of the opening ceremony. The intention is that all students should go to the lectures of invited speakers, even if they are not attending a course at that particular time.

*ELSN*ews is published at the School of Cognitive and Computing Sciences, University of Sussex. It is printed by the University of Sussex Print Unit.

Editor: Jenny Norris.

Editorial Team: Geoffrey Sampson, Steven Krauwer and Brigitte Burger.

ISSN 1350-990X

© ELSNET 2001

FOR INFORMATION  
Contributions to  
*ELSN*ews, and corrections  
should be sent to:  
jennyn@cogs.susx.ac.uk

Tel: +44 1273 678659  
Fax: +44 1273 671320

Material for the next  
issue is due:  
19 December 2001



### FOR INFORMATION

**Anders Green** is a GSLT graduate student at the Department of Numerical Analysis and Computer Science, Royal Institute of Technology, Sweden

**Email:** green@nada.kth.se

**Web:** www.nada.kth.se/~green/

**GSLT Web site:** www.gslt.hum.gu.se/(en)/

**GSLT Email:** info@gslt.hum.gu.se

**NorFa Web Site:** www.norfa.no



# SYNFACE: A Speech-driven Synthetic Face as a Communication Aid for Hearing-Impaired People

Andrew Faulkner, University College London

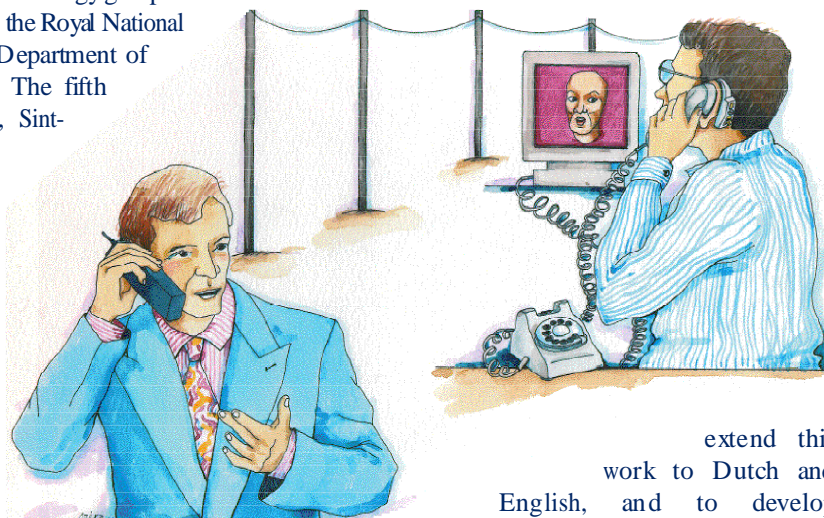
The Framework V IST programme has recently funded a project to develop and evaluate speech-driven facial animation to assist hearing-impaired people in using the telephone. The consortium is led by the Department of Speech, Music, and Hearing at KTH (Royal Institute of Technology), Stockholm. The industry partner is Babel-Infovox AB, a Swedish enterprise that has very recently become part of the Belgian speech technology group BaBel Technologies. The UK partners are the Royal National Institute for Deaf People and the Department of Phonetics and Linguistics, UCL. The fifth partner is Instituut voor Doven, Sint-Michielsgestel, the Netherlands.

Many researchers have found speech and language technologies attractive because of their potential to assist communication by people with disabilities that affect their ability to process or to generate spoken messages. Hearing disabilities are by far the most frequent of these. Some three million people in Western Europe alone suffer from a hearing disability that is sufficiently severe to significantly limit the intelligibility of heard speech. In face-to-face communication, these people are helped by seeing the talker's oral movements, that is, by speech-reading or lip-reading. Communication by telephone, which is increasingly important in work and elsewhere in everyday life, is of course problematic for them.

Some hearing-impaired people are fortunate enough to be able to enlist the help of human lip-speakers who can "interpret" heard telephone speech into speech that can be seen and heard. The aim of the SYNFACE project is to realise a talking face based on a standard PC that can generate useful speech-reading information from telephone speech. In order to construct this non-human visual interpreter, a number of problems need to be solved. Most of the talking heads now appearing for internet and other applications are designed only for those of us with normal hearing; they produce oral movements that may be plausible but are not accurate. Hearing-impaired users of such an application, however, would be very dependent on the accurate delivery of phonetically informative movements.

The SYNFACE project builds on previous work at KTH in the TELEFACE project, which led to a non-real-time system that has already provided significant speech-reading support to hearing-impaired people (Agelfors et

al (1998), in *Proc ICSLP98*, Sydney, Australia; Agelfors et al (1999), in *Proc AVSP'99*, San Francisco, USA). TELEFACE used a hidden Markov model and artificial neural network recognition architecture to identify a phoneme sequence, and a phonetic synthesiser to control the articulation of a texture-mapped 3D wire-frame face model. The tasks of SYNFACE are to



extend this work to Dutch and English, and to develop improved methods of speech recognition and face control. Because of the need to operate during live conversation, speech recognition must be in real-time with a minimum delay. Hence, lexical and higher-order language models cannot be applied. However, statistical modelling of phonotactics at the diphone or triphone level can be exploited. Direct mappings between acoustic and facial movement parameters will also be explored (e.g., Massaro et al (1999), in *Proc AVSP'99*, San Francisco USA). Advances in the control of facial movements and their timing are also planned, along with perceptual studies that will help define the information carried by facial movements and identify the aspects of facial movement whose correlates are crucial targets for the speech recogniser. By 2003 the consortium will be conducting home and workplace trials of a PC-based prototype with hearing-impaired users.

## FOR INFORMATION

Andrew Faulkner is Principal Research Fellow and Research Co-ordinator at the Department of Phonetics and Linguistics at UCL, UK

Email: andyf@phon.ucl.ac.uk

Web: [www.phon.ucl.ac.uk/home/andyf/](http://www.phon.ucl.ac.uk/home/andyf/)

The contact for the SYNFACE project at KTH is Inger Karlsson ([inger@speech.kth.se](mailto:inger@speech.kth.se))

Feature

Autumn  
2001

elsnet  
.....

# RACAI – Romanian Academy Centre for Artificial Intelligence

*Dan Tufis, RACAI, Bucharest*



*Dan Tufis*

*ELSNNews asked Dr Dan Tufis, Director of RACAI, for an update on the HLT activities at the Centre. This is what he had to say ...*

Located in the “House of the Academy”, next to the Parliament Palace, RACAI was established in 1994 to conduct basic research in artificial intelligence (with emphasis on machine learning, natural language processing and conceptual modelling) and to promote scientific cooperation and knowledge dissemination.

knowledge extraction from texts and bi-texts, web technologies, and language-enabled applications (document classification, information retrieval, question answering, intelligent tutoring in CALL). Most of the NLP tools and resources developed at RACAI represent the outcome of 22 international research projects and six national programmes in which it has taken part. RACAI has (co-)organised three international conferences, five international summer schools, and twelve national symposia and workshops.

However, there is another aspect of RACAI’s activity which is less quantifiable but undoubtedly has longer-term impact on the NLP/HLT activities in Romania: raising the policy-makers’ and scientists’ awareness of the need for these activities and of their potential.

NLP research is not new in Romania, but until 1990 it was only represented by very small and isolated groups. Computational linguistics and NLP were practically absent in the curricula of the main Romanian



*Bucharest’s “House of the Academy” – home to the Romanian Academy Centre for Artificial Intelligence*

To begin with some numbers: it has a core of twelve permanent staff (three of whom are Members of the Romanian Academy), seven affiliates (of whom five are scholars from abroad), and a variable number of temporary and part-time staff, mainly MSc or PhD students. For the last five years RACAI has organised hands-on open-access courses on using internet technologies, aimed (mainly) at academic researchers in the humanities. Most researchers at RACAI are involved in NLP/HLT, with special interests in computational mono- and multi-lingual lexicography, language resources (annotated corpora, lexicons, wordnets, grammars), corpus linguistics, parsing,

universities. The idea of creating a centre of the Academy dedicated to AI research and education, “officialising” the domain, sprang up in 1990, immediately after the change of the political regime. General scepticism about the prospects of such an institution meant that it was four years until the proposal finally became reality. Since its establishment, RACAI has had active support from the international community, mainly from EU countries, and from professional associations such as ACL, ELSNET, Union-Latine, and AUPELF-UREF. RACAI has also benefited from the personal involvement of international scholars, and has started an active awareness campaign on

language technology (summer schools, international workshops and conferences, regular tutorials). As a result of this campaigning, 1996 saw one of RACAI's biggest successes, as "language technology" became, for the first time, a distinct research area in the National Research Plan of the Ministry of Science and Technology. Funds became available, and the language and speech "virus" started to spread. Now language and/or speech research groups are to be found in almost any technical university in Romania.

In 1999, together with a group of enthusiastic linguists from the Department of Romanian Language at the University of Bucharest, RACAI established a twinned Centre for Computational Linguistics, launching the first Romanian MSc programme in the area of formal and computational linguistics. As well as providing the faculty staff, RACAI provided the computer infrastructure and network communications facilities. And this year the Alexandru Ioan Cuza University of Iasi launched the second Romanian MSc programme in computational linguistics. The two programmes are closely related, and most of the teaching staff teach on both courses, in spite of the 500 km between Bucharest and Iasi.

The most recent success of RACAI is the establishment of a Commission of the Academy dedicated to language

technologies and resources for Romanian. Supported by institutions from Romania and the Republic of Moldova (linguistic institutes, universities, research units, software companies) and individual scholars from many countries, the Commission is a forum that promotes nationwide synergy in efforts aimed at identifying, harmonising and further developing electronic resources for Romanian. New resources such as a treebank, a reference corpus of contemporary Romanian, a diachronic corpus, several speech databanks, various terminological databanks, and new applications in the language and speech area are among the most important objectives that will be promoted by the Commission. Whether this Commission will be a viable construct remains uncertain. At present there is a lot of optimism, and many hopes and promises. Time will tell...

#### FOR INFORMATION

**Dan Tufis** is Corresponding Member of the Romanian Academy, Director of RACAI, and Head of the NLP Group at RACAI

**Email:** [tufis@racai.ro](mailto:tufis@racai.ro)

**Web:** [www.racai.ro/~tufis](http://www.racai.ro/~tufis)

**RACAI:** [www.racai.ro](http://www.racai.ro)

*continued from page 6*

The panels were among the most lively events, enabling participants to attend public discussions on topics that are sometimes confined to private exchanges. The first panel discussed user needs, the second MT evaluation, the third provided a contextualised view of MT today, and the fourth concerned future developments. The panels also differed in style. In the panel on MT evaluation, for instance, Margaret King asked each of the panelists in turn a series of questions such as "what is the most important thing an evaluator has to bear in mind?", as well as the more traditional "best and worst evaluation". Since the panelists had all been involved in previous evaluations, the discussions exemplified critical points in MT evaluation design.

In another panel, chaired by Laurie Gerber, each panelist had been asked to prepare a short, up-to-date introduction to one of the following topics: the evolution of MT, MT today in America, in Europe, and in South-East Asia. Finally, in his perspective on the future, Eduard Hovy made three interesting predictions: that MT will be included in every web browser and email program; that MT will be available for all language pairs; but that MT quality will remain more or less the same as it is now.

It seems, indeed, that progress in MT quality will depend on a major breakthrough in research. However, it is unclear whether research programmes worldwide do concentrate enough effort on MT. The present MT summit gathered a strong community of technology developers and providers,



*Some participants enjoying the banquet on the Saturday evening*

among which contributions from theoretical research may have seemed scarce. Still, as the panelist mentioned above noted, major advances in MT are conditioned by advances in natural language processing in general, especially in language understanding and in semantic models. It is therefore to be expected that fundamental MT research programmes will encourage research in these directions, and will ease the knowledge transfer towards more efficient MT methods and technologies.

#### FOR INFORMATION

**Andrei Popescu-Belis** is a researcher at ISSCO, ISSCO/TIM/ETI, University of Geneva

**Email:** [Andrei.Popescu-Belis@issco.unige.ch](mailto:Andrei.Popescu-Belis@issco.unige.ch)

**Web:** [www.issco.unige.ch/staff/andrei/andrei.html](http://www.issco.unige.ch/staff/andrei/andrei.html)

**Summit web site:** [www.eamt.org/summitVIII/](http://www.eamt.org/summitVIII/)

**Autumn  
2001**

**elsnet**  
.....



# MT Summit VIII

## A Forum on Machine Translation Technologies

*Andrei Popescu-Belis, ISSCO, University of Geneva, Switzerland*



*Some of the participants enjoying the wonderful setting of the Hostal dos Reis Católicos during a break*

The eighth Machine Translation Summit, from September 18-22, 2001, took place in the beautiful city of Santiago de Compostela, in Galicia, North-Western Spain. Best known as a meeting point for pilgrims from all over Europe, hence a multilingual place, the city today displays a Galician-Castilian bilingual environment. The MT Summit fully benefited from this auspicious setting, which included the magnificent historic building that hosted the conference.

A forum of MT research as well as a showcase of translation and multilingual technology, the summit brought together representatives of universities, companies, and international organisations involved in MT either as developers, providers, or users. The local organisers have kindly compiled a list of all the delegates, providing statistics about their origins. To quote, there were 48% delegates from “industry”, 37% “academic”, 12% “governmental”, and 3% “other”. Overall, European countries accounted for almost two thirds of the participants (with less than 1%, unfortunately, from Central and Eastern European countries), with most of the remaining delegates coming from the United States and South-East Asia. Among European countries, Spain, Germany, and the United Kingdom were the most highly represented.

Significantly, the summit was much more than a scholars’ meeting in which research work is presented and discussed. The organisers – in particular John

Hutchins, conference chair, and Bente Maegaard, programme chair – had designed a schedule that enhanced interactivity, proposing a wide choice of encounter formats, to say nothing of extra-conference events. There were opening and closing sessions, three keynote talks, four panels, a poster session, an exhibition accompanied by nearly ten demos, four workshops, and of course numerous sessions for oral presentations (with three talks per session). The blending of research presentations and technology applications seems thus a successful trend, as illustrated also by the NAACL 2001 conference held this summer and dubbed “Language Technologies 2001”.

It is not possible to summarise here such a rich programme, but an attempt to synthesise the main topics of the twenty-something sessions would point in the first place to the description of applied systems (four sessions), to lexicons and dictionaries for MT (3.5), and to MT evaluation (3). Then come multilingual text generation (2) and the analysis of language-specific techniques (1.5), followed by statistical and corpus-based MT, example-based MT, and alignment plus translation memories (one session each). This is of course only an outline, since presentations often pertained to several subjects. Some of the topics were also to be found in the four associated workshops: example-based MT, roadmap for MT in the next decade, MT evaluation, and teaching MT.

*continued on page 5*

e

# Swords into Ploughshares at Malvern

***ELSNews* Interviews Roger Moore of 20/20 Speech Ltd**

Interview



*Roger Moore*

The elegant English spa town of Malvern is home to what was until 1991 the Royal Signals and Radar Establishment (RSRE), Britain's chief government electronics research organisation. RSRE had a remarkable history of technological innovation – though this often became public knowledge only long after the event, when defence secrecy no longer applied. RSRE was responsible for the World-War-Two-winning invention of radar and had a large role in the development of the digital computer; more recently, liquid crystal display technology was invented at Malvern. In our field, RSRE was home to the internationally-respected Speech Research Unit, led since 1985 by Dr Roger Moore.

British policy over the last twenty years has been, wherever possible, to replace civil service organisational arrangements by the disciplines of the market. For RSRE the first consequence was a merger with other research sites into the (financially autonomous) Defence Evaluation and Research Agency (DERA); and, this year, except for the small part concerned with particularly defence-sensitive work, DERA morphed into QinetiQ plc, a public-private partnership which is Europe's largest science and technology organisation.

But the Speech Research Unit was a step ahead on the privatisation road. In 1999, it became 20/20 Speech Limited, an equity joint venture between NXT plc (as majority shareholder) and DERA. 20/20 Speech occupies smart new premises at the Malvern Hills Science Park, adjacent to the old RSRE site but, significantly, outside the perimeter wire guarding Britain's military secrets.

"Marketisation" is a theme of our times; doubtless many researchers elsewhere in Europe will be treading paths similar to 20/20 Speech's in years to come. *ELSNews* invited Roger Moore to talk about his group's transition from scientific civil servants to private-sector researchers.

***ELSNews:*** Who are NXT plc and why did they want to acquire you?

***Moore:*** NXT was formed out of the old Verity group of UK-based hi-fi manufacturers – Mission, Wharfedale and Quad were three of their brands, for instance. They transformed their business as a consequence of exploiting a DERA patent. Ten years ago, Ken Heron of DERA was working on eliminating noise from vibrating panels to make life in military aircraft more bearable, and he realised that if he turned his equations upside down they could make a flat surface act as a speaker – NXT's subsequent research into distributed mode vibrations led to the invention of what they now call SurfaceSound and SoundVu. Because the sound is not produced pistonically from a point source, many traditional problems of speaker technology fall away. You can use a television or projector screen as a speaker. You can incorporate fire alarms into fire-retardant tiles. The speaker of a home hi-fi system can be a work of art hanging on the wall.

The possibilities are endless, so rather than trying to exploit them all itself, NXT decided to focus on licensing technology worldwide rather than manufacturing hi-fi. NXT was formed in 1998, two years after the group originally launched the flat panel technology and three years after the licence was taken up. Its next step was to find new sound-related technology to license, so speech was an obvious avenue for NXT to explore. Our Unit offered their best opportunity in this respect, and for us NXT represented our most promising path into the future.

***ELSNews:*** Why were your group keen to move into the private sector?

**Autumn  
2001**

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



**Moore:** In civil service days, our funding was by a block grant governed by agreed headcount. The good aspect was great freedom to decide for ourselves which research avenues to explore. But the downside was no growth possibilities. In fact, with the end of the Cold War, headcounts were being reduced; the fact that our technology had burgeoning civilian applications could not be used to offset that trend. Our speech-recognition engine, for instance, was state-of-the-art; so far as the technology is concerned it might have been the basis for the kind of voice-input system that Dragon and IBM have brought to market very successfully. In practice, that could not happen. For a government agency to take equity in a private company is an arrangement that only became legally possible in 1999, a few months before 20/20 was founded.

Each member of the Speech Research Unit was individually given the choice of joining 20/20 or staying in DERA. Every one opted to make the move.

The very fact of being “behind the wire” was itself a drag on development. True, some industry visitors would think “If their work is this well guarded, it must be important”. But overall it was a psychological as well as organisational barrier. We now know *what* we are doing, *for whom*, and *why*. Incidentally, over the last year we have taken out more patents than in the previous ten years put together.

**ELSNNews:** Even though the nature of your work was what attracted NXT, within 20/20 Speech it must have changed in some ways?

**Moore:** For sure. In the first place, we used to be very wide-ranging – there were few branches of speech research where we were not active. For 20/20 Speech, that strategy wouldn't work; we necessarily focus on a limited number of lines with market potential. But then we always worked in a collaborative fashion, with each member contributing to several different projects, so nobody identified themselves individually with one particular area that was dropped.

A bigger change has been the re-orientation towards product development. Within DERA our work was about “proof of concept”: if we could show that something was possible, it was not our job to turn it into a saleable product. In 20/20 Speech we have brought in new staff with expertise in industrial software engineering, marketing and so on, but with advanced technology like ours it would be unrealistic for the scientists to stand back and leave “productisation” entirely to people with a commercial background; so most of the speech experts have got drawn into that side of things.

We used to participate actively in academic conferences and publication. In principle, we still can; but it remains to be seen how well that will combine with our new work priorities in practice.

The pressures are very different. All of us agree that you can never appreciate this fully, until you live it. But I don't think any of us would choose to go back.

**ELSNNews:** What are the 20/20 Speech development priorities?

**Moore:** We are focusing on two themes:

- People on the Move
- Speech in Media

“People on the Move” includes gearing applications to eyes-busy/hands-busy situations – for instance, allowing drivers to operate some controls by voice – and creating speech applications with a small footprint, usable in cramped conditions.

“Speech in Media” includes such things as computer-assisted subtitling (adding subtitles to a television programme manually takes at least sixteen hours per hour of programme – currently we can reduce that by 30-50 per cent), and automating various aspects of studio operation, for instance, voice-driven teleprompts, and synchronising dubbed voices.

Some of our work is still done for the Ministry of Defence. But, now, they are just one client among others. At the other extreme, some of it relates to pure entertainment, for instance, improving the realism of computer games.

**ELSNNews:** Good luck with the new venture! We wish you well.

#### FOR INFORMATION

**Roger Moore** is Chief Scientific Officer at 20/20 Speech Ltd. He is visiting Professor in the Department of Phonetics and Linguistics, University College London, and served as the President of the International Speech Communication Association (ISCA) from 1997 to 2001

**Email:** R.Moore@2020speech.com

For more information about 20/20 Speech Ltd, visit their **Web site:** [www.2020speech.com/](http://www.2020speech.com/)

e



# SIGdial: Stray Thoughts from a Crossroads

*Hans Dybkjær, Prolog Development Center A/S, Brøndby, Denmark*

*Hans Dybkjær gives a commercial-eye-view of the 2nd SIGdial Workshop, held on September 1-2, 2001 in conjunction with Eurospeech 2001.*

This year's SIGdial workshop took place in Aalborg, Denmark, and I must say that it was very well organised.

A great thing about dialogue and discourse is that, on the one hand, it is deeply theoretical and philosophical, and on the other, it addresses practical, commercial-level problems of satisfying system users. Looking at the audience of this workshop, you see both theorists and engineers, and you know that over the years they will move back and forth across the boundaries of academia and industry.

I'm one of them. Today I am "out there", building commercial information-service systems using the matured forms of the kind of tools and platforms we had in research five to ten years ago.

So, what am I after? Well, I need research and researchers to provide me with wider insights, concepts, and methodologies: hard evidence to steer my work. In this respect, the rhetorical structure theory tagging experiments of Carlson et al, are close to perfect: the theory itself seems conceptually appealing, and the careful investigation of the validity of this by extensive intercoder reliability experiments is something everybody should learn from. Though we still fail to see dialogue control tools reflecting RST, this kind of experiment tells me that I can robustly rely on insights gained from RST.

As a working practitioner I am drawn to everything that tells me I can do things – and in a simple way – and to everything that describes the problems I can expect to meet. And advice needs to be given by people who approach problems at the same level as I do (although that is not to imply that I can solve them as elegantly). From this perspective I loved the reports from the Spanish railway information systems by San-Segundo et al, providing, for example, a cookbook on graceful degradation. And the dauntless practical systems presented by Rudnicky are just pure gold. But I also listened carefully to empirical work like Cavazza's recognition errors or to Aberdeen's reports on the hopeless users in their experiments.

In my daily work, I not only use state-of-the-art tools, I also push them to the limits of their intended use. So, when Hirschberg convincingly told us that prosodic cues today can point out the correction awareness sites in practice, I immediately asked Philips (because I use their otherwise nice SpeechMania platform) if they

could kindly provide me with that technology. And I am mathematically inclined to love the turn-minimisation criteria described by Yasuda et al, and to believe they could be incorporated into a dialogue control tool: so why aren't they there?

This brings me to the issue of crossing the gap from theory to practice. The turn-minimisation is formalised and can be made operational. Don't forget that at the bottom of software engineering lies programming, which is nothing but applied, extremely formal logic. But what about the prosodic cues? Researchers may need to create the tools, not only as a necessary part of the research projects, but also to convince companies about the practical applicability of their work. And to mature the frameworks: the "plug'n'play" ideas for grammars propounded by Lewin et al sound attractive and are, engineering-wise, desirable. I am, though, not convinced that they have got the abstractions right – the interfacing of their grammars makes me suspicious.

In general I will not use research-produced software. I love to try it out, and also enjoyed doing that at the workshop. However, at the end of the day I use software that has been made by professionals who know about basic engineering skills like regression tests and version control, and who are geared to support.

Summing up, the mix provided by this year's SIGdial workshop was great, although I might have wished to see some more hard and weird theory – and more of the theory subjected to practical and empirical scrutiny.

But at the end I just want to say, to the organisers and not least to the contributors: "Keep up the good work!"



SIGdial  
(ACL Special  
Interest group on  
Discourse and  
Dialogue)

## FOR INFORMATION

**Hans Dybkjær** is software developer and project leader at the Prolog Development Center A/S

**Email:** dybkjaer@pdc.dk

**SIGdial Web site:** <http://www.sigdial.org/>

The proceedings of the workshop, which include all papers referred to in the text, can be obtained in print from the ACL, and are also available from the SIGdial web site at [www.sigdial.org/sigdialworkshop01/](http://www.sigdial.org/sigdialworkshop01/)

**Autumn  
2001**

**elsnet**  
.....

# BalkaNet

## A Multilingual Semantic Network for the Balkan Languages

*K. Oflazer, Sabanci University, Turkey; S. Stamou and D. Christodoulakis, Patras University, Greece*



*Kemal Oflazer*

The BalkaNet project now being funded by the EC for three years under the IST programme aims at building a multilingual lexical database consisting of wordnets in the following languages: Greek, Turkish, Romanian, Bulgarian, Czech, and Serbian. Members of the BalkaNet consortium are: University of Patras (Greece);

Computer Technology Institute (Greece); Alexandru Ioan Cuza University of Iasi (Romania); Centre for Advanced Research in Machine Learning, Romanian Academy (Romania); Bulgarian Academy of Science (Bulgaria); Sabanci University (Turkey); Faculty of Informatics, Masaryk University (Czech Republic); Memodata (France); University of Plovdiv (Bulgaria); and University of Athens (Greece). The subcontractor sites will be: Faculty of Mathematics, University of Belgrade (Yugoslavia); Centre of Applied Research (the Netherlands); and the Internet Service Provider, Otenet (Greece).

The success of WordNet (Miller et al, 1990) has motivated similar projects that aim at the development of wordnets for languages other than English. A European project called EuroWordNet (EWN) (Vossen, 1998) added eight other languages to this thesaurus, resulting in a huge network of linguistic concepts that allow inter-lingual navigation and search of translation equivalences between languages via a so-called Inter-Lingual-Index (ILI). The latter originated from WordNet 1.5 synsets and has been restructured on the basis of aspects of the internal structure of the monolingual WordNets. The characteristics of the ILI are defined by its function to provide an efficient mapping across the meanings in the wordnets for the different languages (Vossen, 1999).

The BalkaNet project aims to develop a multilingual lexical resource representing semantic relations among basic concepts of the Balkan languages. The Balkan wordnets will be built as far as possible from available lexical resources, so that it will be possible to combine information from independently created resources, making the final database more consistent and reliable, whilst at the same time retaining the richness and diversity of the vocabularies of the languages involved. The main resources of information will be the individual monolingual wordnets

that have already been developed or are currently under development for most of the participant languages. Where a monolingual wordnet is not available, dictionaries, thesauri, or corpora of the respective languages will be used for the terminology extraction.

Within the framework of the BalkaNet project we aim at a total set of 15,000 comparable synsets in each language, corresponding to more or less 30,000 distinct word forms, covering generic vocabulary of the languages involved. The part-of-speech (POS) distribution will be 65% nouns, 25% verbs, 5% adjectives, and 5% adverbs. In addition, the monolingual wordnets developed from scratch within the framework of the project will comprise approximately 8,000 synsets, whereas the number of synsets that will be added in already existing wordnets will be determined at a later stage. In addition, apart from the representation of generic vocabulary in the multilingual database, a feasibility study will take place in order to test how domain-specific terminology can be incorporated into the semantic network under domain labels. Finally, the BalkaNet database will be incorporated into the EWN semantic network, resulting in a global semantic database covering conceptual areas of European languages.

The BalkaNet semantic network can be used directly in many NLP applications, ranging from word sense disambiguation tasks to language-learning tools and dictionary publishing. In addition, the multilingual database may serve as a starting point for large lexical knowledge bases or as a source of semantic information to improve grammar- and spell-checkers.

One envisaged application of BalkaNet concerns its incorporation in information retrieval (IR) systems in order to support conceptual text retrieval as opposed to exact keyword matching. The available linguistic tool which will support textual research will not be restricted to the English language, thus facilitating access to the stored data for non-English native speakers and making multilingual IR feasible.

The most immediate application of the BalkaNet in IR tasks concerns expanding the query by including members from semantically related senses of the original query term, thus improving recall of the obtained results. The hierarchical structure of the thesaurus enables broadening the search with synonyms and other related words, creating a basis for more accurate and relevant information retrieval performance. Moreover, the BalkaNet semantic network can be applied in (word)

sense disambiguation tasks as well. Despite the fact that WordNet itself does not give any information about the context in which a wordform or its senses occur, nevertheless (as pointed out by Miller in Miller & Charles (1991)) knowing how a term interacts with other terms in a context is a requirement for knowing a term. In general, there are additional applications where a multilingual semantic network can be used, such as in machine translation, semantic annotation, and so on.

### Acknowledgements

BalkaNet research is supported by the European Commission in the framework of the IST Programme (IST-2000-29388). We wish to thank all partners of the project for their valuable contributions, and Dr. Piek Vossen and Prof. Cristiane Fellbaum for their support.

We also thank the other members of the consortium: Karel Pala, Dan Cristea, Dan Tufis, Svetla Koeva, George Totkov, Dominique Dutoit, and Maria Grigoriadou.

### References

Miller, G.A., Beckwith R., Fellbaum C., Gross D., Miller K.J. (1990) Introduction to WordNet: An On-Line Lexical Database. In *International Journal of Lexicography*, Vol. 3, No. 4, pp. 235-244

Miller, G.A. & Charles, W.G. (1991) Contextual correlates of semantic similarity. In *Language and Cognitive Processes*, pp. 1-28

Vossen, P. (1998) *A Multilingual Database with Lexical Networks*. Kluwer. Academic, Dordrecht.

Vossen, P., Peters W., & Conzalo J. (1999) Towards a Universal Index of Meaning. In *Proceedings of the ACL'99 Siglex Workshop*, University of Maryland

### FOR INFORMATION

**Kemal Oflazer** is a faculty member in the Human Language Technology Laboratory at Sabanci University in Istanbul

**Email:** oflazer@sabanciuniv.edu

**Web:** www.sabanciuniv.edu/fens/people/oflazer/

**Sofia Stamou** and **Dimitris Christodoulakis** are both in the Department of Computer Engineering & Informatics at Patras University

**Email:** stamou@cti.gr; dxri@cti.gr

**BalkaNet Web Site:** www.ceid.upatras.gr/Balkanet  
Please note that this site is still under construction

# Lexicom 2001: Masterclass Workshop on Lexicography and Lexical Computing

*Sue Atkins, Information Technology Research Institute, University of Brighton*

“Learning how to” rather than “learning about” was the keynote theme of the first Lexicom Workshop on Lexicography and Lexical Computing which took place at the Information Technology Research Institute, University of Brighton, England, from July 16-20, 2001. The focus of the tuition and practical work was twofold: the making of corpus-based dictionaries, and the use of software that supports this activity.

### Participants

Fifty participants from 21 countries spent a week of intensive training, partly in seminars and partly in hands-on practice in the computer suite, where everyone had their own machine and could choose to work on the practical tasks either individually or in informal small groups.

Members of the first Lexicom came from many different backgrounds: publishing houses (managers, senior editors, and junior editors); universities (linguists, computational linguists, computer scientists, academic lexicographers, and graduate students); software houses (software engineers and computational linguists); government agencies (terminologists and translators);

and various types of research institution. They formed a lively and enthusiastic group, sharing their ideas and experience, and always ready to join in the discussions and keep the tutors on their toes

### Programme

Our aim was to give participants a grounding in both lexicography and lexical computing. They had been accepted on the basis of experience in one of the foundation disciplines: lexicography, computer science, and linguistics, and of course everyone had something to offer and everyone learned from others on the course. The experience of learning among friends was one of the most distinctive features of the workshop.

The programme is to be found in detail on the website. The two types of activity – seminars and practical work – reflected the essentially practical nature of the workshop: each seminar introduced and discussed a key topic, which was then further explored in practical exercises at the computer. As well as the basics of practical lexicography, from the structure of the dictionary entry through corpus querying (using the

Masterclass  
Workshop Report

**Autumn  
2001**

**elsnet**  
.....



WordSmith Tools software on the British National Corpus) to analysis of the data and compiling actual dictionary entries, the programme included an overview of: corpus design and annotation and the extraction of information from corpus data; the use of sophisticated data-mining software (Kilgarriff's "Word Sketches") for providing lexicographers with profiles of a word's salient combinatory features; a brief look at various types of dictionary databases; and an introduction to frame semantics as a basic and practical approach to corpus analysis. Participants received complete documentation of all the seminars in a bound copy of the Course Notes.

There were also opportunities for participants to present their own work to the group: one session consisted of presentations by six participants of their own software which is already in use – a rewarding experience for us all. And in an enlightening and

computational work. It was hard work but it was great fun, and we all learned a lot. Thank you to everyone who came to work with us – you were a great bunch of folks!

Participants' comments on the workshop can be found on the website.

### Lexicom 2002

This will be held from 14 to 19 July. Like its predecessor, it will be a workshop aimed at practitioners learning and practising the basics of their own and related disciplines. The programme and general approach for 2002 will be substantially the same, with some enhancements as suggested by this year's group. Pre-registration is already taking place on the website. The dozen or so people who unfortunately didn't get further than the waiting list for Lexicom 2001 will be the first to be offered places on the next course.



*The participants of Lexicom 2001: Sue Atkins is in the front row, third from left; Adam Kilgarriff is at the far right in the back row; Michael Rundell is at the very back, five rows behind Sue Atkins*

enjoyable closing session, several groups presented dictionary entries they had compiled, initiating discussions of various theoretical and methodological issues and of the respective contributions of the corpus-query software, Word Sketches, and material from the FrameNet database.

As for social events, we all met at the opening reception and later in the week got together for a cheap and cheerful fish and chip supper on Brighton Pier. Brighton proved an ideal setting for spare time activities, from walks along the beach to an evening's clubbing for those with real stamina.

### Tutors

The course was led by Sue Atkins, Adam Kilgarriff, and Michael Rundell, the partners in the Lexicography MasterClass who together can claim about seventy years' experience of dictionary-making and dictionary-oriented

### FOR INFORMATION

**Sue Atkins** has been an Honorary Visiting Fellow of ITRI since 1998, and received an Hon. D. Litt from the University of Brighton in July 2000 for services to lexicography.

**Email:** [btsa@itri.brighton.ac.uk](mailto:btsa@itri.brighton.ac.uk)

**Email** for queries about the workshops: [Lexicom@itri2001](mailto:Lexicom@itri2001)

**Lexicom home page:** [www.itri.brighton.ac.uk/lexicom/index.html](http://www.itri.brighton.ac.uk/lexicom/index.html)

**Pre-registration for Lexicom 2002:**  
[www.itri.brighton.ac.uk/lexicom/reg\\_form.html](http://www.itri.brighton.ac.uk/lexicom/reg_form.html)

Information about the **Masterclass Partners:**  
[www.lexmasterclass.com](http://www.lexmasterclass.com)

Announcement



# Bean Counting

*John Nerbonne, University of Groningen*

Opinion Column

Lots of us deal with the “publish or perish” demand. Those working at universities and research institutes are expected to publish regularly, and larger companies are often pleased to see the output of their research laboratories measured not only in sales of new product lines but also in the number and quality of publications.

The requirement is reasonable. First, research that isn’t communicated in some accessible channel hasn’t added usefully to what we know: it has to be published. Second, even if the primary task of universities is education, as it was until fifty years or so ago, still educators need be intellectually active, and the requirement of publication provides some objective reassurance to educational administrators that their staff are not intellectual zombies. Third, and most fundamentally, publication is part of the scientific dialectic: by publishing your findings, you’re exposing them to the criticism that will identify flaws, gaps, inconsistencies, and gratuitous assumptions. The mistakes that the next paper – yours or mine – should remedy. Publish!

One hears the complaint that too much is published – that it’s not all worth reading and that is true in some sense. No one wants to argue for lower standards. But this just means that we ought to be more selective about quality work, not give up publishing.

The requirement to publish is operationalised more ambitiously, not just as a demand to publish, but as a system of evaluation to measure research quality via publication. For someone engaged in the language technology of the last ten years it only seems fair that one try to fix the evaluation measure somehow. First, not just any publication will do. Journals count more than books, book chapters, and conference proceedings. I have a grant from one European organisation that requests reports of publications, but only those in international, refereed journals. Book chapters and papers in COLING or ACL proceedings simply aren’t worth mentioning. Second, not all journals are equal. Journals with high impact ratings as measured by the *Science Citation Index* (SCI, see [www.isinet.com](http://www.isinet.com)) count more heavily. The SCI rates a journal by how often other journals cite its articles

Both of these principles have some initial plausibility. We all know that publication channels differ, and the review process at journals is arguably more reliable than the process for book chapters and conference proceedings. And journals certainly differ in quality.

But lots of qualifications are needed. In computational linguistics, competition for slots at the leading meetings results in a lower acceptance rate (typically around 25-30%) than many leading journals (the acceptance rate for *Computational Linguistics* is only slightly more selective,

20-25%). There is a tradition of strict selection that will be difficult to defend if research funders systematically discount – even disregard – conference contributions. This tradition promotes the quality of conference presentations, and we’ll lose something if it becomes weakened.

The use of citation indices likewise has an initial plausibility that is subject to abuse. It is our task as scientists to find out new things and to change our colleague’s minds about how best to understand language and computation. The number of citations certainly reflects that better than other measures, say, the number of pages produced (the measure used in my faculty until recently). But there’s many a slip twixt cup and lip in tracking citations. The SCI doesn’t include citations in journals such as: *Journal of Logic, Language and Information*; *Journal of Natural Language Engineering*; *Journal of Functional Programming*; *Journal of Computer Languages*; *Computer-Assisted Language Learning*; or *Traitement Automatique des Langues* (check the “Master Journal List”, accessible from [www.isinet.com/isi/search/](http://www.isinet.com/isi/search/)). It now tracks citations in about 8,000 of the world’s approximately 15,000 scholarly journals (thanks to Groningen’s university librarian, Alex Klugkist, for this latter figure). I noted the examples above when trying to understand how a move to using the SCI measurements would affect the assessment of our computational linguistics group in Groningen, but I’ve left more specialised and Dutch-language journals off this list.

A further source of distortion is that articles in specialised journals are generally rated as less important, since these journals attract fewer citations overall. An authorial strategy of seeking out the most general venue (ultimately *Science* or *Nature*) is rewarded even when the most expert reviewing and selection would be found elsewhere. The same frequency weighting will inevitably distort comparisons between larger disciplines and smaller ones (say, chemistry versus computational linguistics), but this doesn’t prevent the ratings from being abused in order to compare across disciplines.

Web publication is certainly going to rationalise the distribution of scientific literature, but it won’t obviate the need for systems for selecting the better papers (refereeing), nor has anyone proposed how it might change the socioeconomic, political question of choosing which researchers and which sorts of research deserve funding.

## FOR INFORMATION

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

**Email:** [JNerbonne@let.rug.nl](mailto:JNerbonne@let.rug.nl)

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

**Autumn  
2001**

**elsnet**  
.....

# Just a little Imagination at Eurospeech 2001

*Gerrit Bloothoof, Utrecht University*

In *ELSNews* 9.3 we announced **Imagination 2001** – a competition for young and innovative researchers in the field of *HLT*, to be held in association with Eurospeech 2001, with a prize of 5000 Euro (see Editor's Note below) for the most imaginative application. Gerrit Bloothoof reports ...

At Eurospeech 2001 in Aalborg, ELSNET challenged the young generation to demonstrate or simulate creative applications of language and speech technology. Hewlett-Packard European Research Labs supported the contest with a prize of 5000 Euro. Despite this financial incentive, only seven young researchers demonstrated a series of interesting applications, which nevertheless attracted many of the participants of the conference.

Multimodal web and personal assistants were a popular theme, but there was also an aid for blind programmers, a graphical tool for dialogue system design, and an interactive introduction to sign language. The jury (Furui, Mariani, Hirshberg, Tucker) concluded that none of the contribution scored high on all their criteria and divided the price between two winners:

- João Valente for his contribution with Arminda Guerra entitled *Gestural Language Introduction*
- Tao Jianhua for his contribution with Cai Lianhong, Wu Zhiyong, and Wang Zhiming entitled *Wise Assistant and Voice Gateway Based on Chinese TTS system SinoSonic*

Abstracts of all the contributions can be found at the website: [www.elsnet.org/imagination2001](http://www.elsnet.org/imagination2001)

## FOR INFORMATION

**Gerrit Bloothoof** is a researcher and lecturer at the Utrecht Institute of Linguistics OTS in The Netherlands. He is the convenor of ELSNET's *Training Task Group*, and is organiser of the Imagination 2001 event.

**Email:** [gerrit.bloothoof@let.uu.nl](mailto:gerrit.bloothoof@let.uu.nl)

**Web:** [www.let.uu.nl/~Gerrit.Bloothoof/personal/](http://www.let.uu.nl/~Gerrit.Bloothoof/personal/)

## Editor's Note

*ELSNews* would like apologise for the lack of the symbol for the Euro – a highly relevant and embarrassing omission for a publication with such wide European coverage.

We would like to offer a brief explanation for this shortcoming, that will have been noticed in past issues by observant readers.

The software used to produce *ELSNews* is QuarkXPress, and in all respects other than this it has proved its worth completely.

With the increasing requirement for the Euro symbol, *ELSNews* has tried to obtain it from the makers of QuarkXPress, but has been advised that this can only be done by purchasing a new version of the international software package. Since the cost of this is prohibitive (some 2000 Euro), we can only ask our readers to accept the substitution of the word for the symbol – in the hope that the matter will be resolved in the not-too-distant future.

If any of our readers have a solution to this problem we would be very grateful to hear from them.

## Call for Letters

At *ELSNews* we welcome letters from readers about any subject related to the area of Human Language Technologies. Thus, if you have any strong views about anything you have read in past issues of the Newsletter, or would just like to express your opinion about an HLT-related matter, then do send in your letters to the Editor, Jenny Norris

**Email:** [jennyn@cogssussex.ac.uk](mailto:jennyn@cogssussex.ac.uk)

Please remember that the deadline for the next issue is December 19, 2001



## Future Events in 2001

- Nov 29-30**      *23rd Annual Conference on Translating and the Computer*, London, UK.  
Email: nicole.adamides@aslib.co.uk; URL: www.aslib.co.uk
- Nov 30**      *12th Meeting of Computational Linguistics in the Netherlands (CLIN 2001)*: Twente, The Netherlands.  
Email: clin@cs.utwente.nl; URL: parlevink.cs.utwente.nl/Conferences/clin2001.html
- Dec 3-5**      *Information Society Technologies (IST 2001) – Technologies Serving People* (European Commission Event):  
Düsseldorf, Germany. Email: info-ist2001@cec.eu.int  
URL: 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: asru01.itc.it
- Dec 11-13**      *IRCS Workshop on Linguistic Databases*: Philadelphia, USA.  
Email: sb@ldc.upenn.edu URL: www.ldc.upenn.edu/annotation/database/
- Dec 14-15**      *International Workshop on Information Presentation and Natural Multimodal Dialogue*: Verona, Italy.  
Email: class-ws@itc.it URL: i3p-class.itc.it/events/ipnmd2001.html
- Dec 17-19**      *13th Amsterdam Colloquium (AC2001)*: Amsterdam, The Netherlands.  
Email: ac2001@wins.uva.nl URL: www.illc.uva.nl/AC2001/

## Events Coming in 2002

- Jan 3-7**      *Literature Data Mining for Biology* (special session within the *Pacific Symposium on Biocomputing 2002*):  
Lihue, Hawaii. Email: russ.altman@stanford.edu URL: psb.stanford.edu
- Jan 7-18**      *LOT Winter School 2002*: Leiden, The Netherlands.  
Email: lot@let.uu.nl URL: www.lot.let.uu.nl/GraduateProgram/graduate.htm
- Jan 8-9**      *5th Annual CLUK Research Colloquium*: Leeds, UK.  
Email: jre@comp.leeds.ac.uk URL: www.dcs.shef.ac.uk/research/cluk
- Jan 21-25**      *1st International WordNet Conference*: Mysore, India.  
Email: secretary\_gwn@ciil.stpmysoft.net URL: http://www.ciil.org/gwn
- March 24-27**      *Human Language Technology Conference (HLT 2002)*: San Diego, California, USA.  
URL: hlt2002.org
- March 25-26**      *Speech Technology in the Learning and Assistive Interface Symposium (InSTIL 2002)*: San Diego, California, USA.  
Email: p.delloque@msec.acuk URL: dbstay.acuk/instil
- April 8-10**      *ISCA International Tutorial and Research Workshop on Temporal Integration in the Perception of Speech (TIPS)*:  
Aix-en-Provence, France.  
Email: tips@lpl.univ-aix.fr URL: www.lpl.univ-aix.fr/%7ETips
- April 8-10**      *The Processing of Arabic*: Tunis, Tunisia.  
Email: brahem@isit.rnrt.tn
- April 24-25**      *Language Technology for Business Information Systems – Special Session in conjunction with the 5th International  
Conference on Business Information Systems (BIS 2002)*: Poznan, Poland.  
Email: piskorsk@dfki.de URL: bis.kie.aepoznan.pl

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

**Autumn  
2001**

**elsnet**  
.....

## ELSNET

## Office

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

## Task Groups

*Training & Mobility*  
Gerrit Bloothoof,  
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,  
LIMS-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,  
LIMS-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 Vienna University of Technology  
A University of Vienna  
B University of Antwerp -UIA  
B Katholieke Universiteit Leuven  
BG Bulg. Acad.Sci. - Institute of Mathematics  
and Informatics  
BY Belorussian Academy of Sciences  
CH SUPSI University of Applied Sciences  
CH University of Geneva  
CZ Charles University  
D Universität Stuttgart-IMS  
D Christian-Albrechts University, Kiel  
D Universität Hamburg  
D Institut für Angewandte Informationsforschung  
D German Research Center for Artificial  
Intelligence (DFKI)  
D Universitaet des Saarlandes  
D Universität Erlangen-Nürnberg -FORWISS  
D Ruhr-Universität Bochum  
D Universität des Saarlandes CS-AI  
DK Center for Sprogteknologi  
DK University of Southern Denmark  
DK Aalborg University  
E University of Granada  
E Universitat Autònoma de Barcelona  
E Universidad Nacional de Educación a  
Distancia (UNED)  
E Universidad Politécnica Madrid  
E Universidad Politécnica de Catalonia  
E Universidad Politécnica de Valencia  
EL Institute for Language & Speech  
Processing (ILSP), Athens  
EL National Centre for Scientific Research  
(NCSR) 'Demokritos', Athens  
EL University of Patras  
F LIMS-CNRS, Orsay  
F LORIA, Nancy  
F Université Paul Sabatier (Toulouse III)  
F Université de Provence  
F Inst. National Polytechnique de Grenoble  
F IRISA/ENSSAT, Lannion  
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 Università degli Studi di Pisa  
I IRST, Trento  
I Consorzio Pisa Ricerche  
I Fondazione Ugo Bordoni  
IRL Trinity College, University of Dublin  
IRL University College Dublin  
LT Inst. of Mathematics & Informatics  
NL Utrecht University  
NL University of Amsterdam (UvA)  
NL University of Nijmegen  
NL Foundation for Speech Technology  
NL Tilburg University  
NL University of Twente  
NL Eindhoven University of Technology  
NL Universiteit Leiden  
NL Netherlands Organization for Applied  
Scientific Research TNO  
NL University of Groningen  
NO University of Bergen  
NO Norwegian University of Science and  
Technology  
P University of Lisbon  
P New University of Lisbon  
P INESC ID, 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 University of Cambridge  
UK Leeds University  
UK University of Sunderland  
UK University of Cambridge  
UK University of Sheffield  
UK University of York  
UK UMIST, Manchester  
UK University of Essex  
UK University of Edinburgh  
UK University of Sussex  
UK University of Brighton  
UK University College London  
UK University of Ulster  
UK University of Dundee

## Industrial Sites

B Lemout & Hauspie Speech Products  
D Philips Research Laboratories

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

## What is ELSNET?

ELSNET is the European Network of Excellence in Human Language Technologies. ELSNET is sponsored by the Human Language Technologies programme of the European Commission; its main objective is to foster the human language technologies on a broad front, creating a platform which bridges the gap between the natural language and speech communities, and the gap between academia and industry.

ELSNET operates in an international context across discipline boundaries and deals with all aspects of human communication research which have a link with language and speech. Members include public and private research institutions and commercial companies involved in language and speech technology.

ELSNET aims to encourage and support fruitful collaboration between Europe's key players in research, development, integration, and deployment across the field of language and speech technology and neighbouring areas.

ELSNET seeks to develop an environment which allows optimal exploitation of the available human and intellectual resources in order to advance the field. To this end, the Network has established an infrastructure for the sharing of knowledge, resources, problems, and solutions across the language and speech communities, and serving both academic

and industry. It has developed various structures (committees, special interest groups), events (summer schools, workshops), and services (website, e-mail lists, \_ELSNews\_, information dissemination, knowledge brokerage).

## Electronic Mailing List

elsnet-list is ELSNET's electronic mailing list. E-mail sent to [elsnet-list@let.uu.nl](mailto: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](mailto:elsnet@let.uu.nl).

## Subscriptions

Subscriptions to *ELSNews* are currently free of charge. To subscribe, 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](mailto:elsnet@elsnet.org)

**Web:** <http://www.elsnet.org>

e