

COLING goes to the Orient!

This year's COLING took place in Taiwan. From the reports it sounds as though future COLING organisers could take a few tips.

*Robert Dale, Macquarie University
Pictures courtesy of Sebastian Varges,
University of Edinburgh*

For many in Taipei, the preferred method of transport would appear to be two wheels. The streets are filled with motor scooters. There are so many that even experienced jaywalkers much prefer to use the traffic-lighted road crossing points. The designers of these are well aware of the fragile coexistence of pedestrian and motorised pedaller: the Taiwanese equivalent of a Walk/Don't Walk sign is an animated little green stick man who walks faster and faster as the countdown clock above him decrements the seconds you have left to escape, if not certain death, an array of cross-body tattoos designed by Michelin or Good year.

COLING 2002, held in Taipei, Taiwan, attracted some 500+ participants to a conference that included two days of workshops, four days of main session papers with a day excursion in the middle, and two days of post-conference workshops. All of this made for an exciting conference, with a particularly large attendance and participation from countries in Asia; hardly surprising given the location, but a great opportunity to learn more about a wide variety of work in computational linguistics that many of us in the west have a tendency to ignore.

With so many papers on offer (over 200 in the main conference alone, spread across five parallel sessions), it would be pointless to pick out individual



The computational linguist's preferred mode of transport

papers for special mention; too often I had irresolvable conflicts in determining what to attend. For me, though, the highlights of the conference were the three invited talks. The first of these, 'Computer Modelling of Language Evolution' by Professor William Wang, was a masterful and fascinating presentation giving an overview of his work on language change across the last 30 years: an excellent opener for the conference. The second invited talk, 'New Chances for Deep Linguistic Processing' by Hans Uszkoreit, was an appeal to consider the proper relationship of shallow and deep processing, and an exhortation not to ignore, in our headlong rush to find shallow solutions for everything, what we have developed in the last 40

Inside this Issue

| | |
|--|----|
| COLING goes East! Robert Dale | 1 |
| Workable Efficient Language Documentation: a report and a Vision Dafydd Gibbon | 3 |
| Text mining as a major advancement for Knowledge management Jon Herring | 6 |
| SIGdial: SmartKom Wolfgang Minker | 8 |
| EDILOG: Dialogue and dancing Paul Piwek | 9 |
| UNISYN lexicon release Lynne Cahill | 10 |
| Endangered Languages Press Release | 11 |
| Opinion Column Kenneth Church | 13 |
| Future Events | 15 |

**Autumn
2002**

elsnet



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: Lynne Cahill.

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

ISSN 1350-990X

© ELSNET 2002

FOR INFORMATION
Contributions to
ELSN_ews, and corrections,
should be sent to:
lynneca@cogsusx.ac.uk

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

Material for the next
issue is due:
13th December 2002



Martin Kay,, Winfried Lenders, and Nicoletta Calzolari at the banquet

years. The third invited talk, on the last day of the main conference, was a very comprehensive and informative presentation on the FrameNet project by Charles Fillmore.

Almost as numerous as the motor scooters were the orange T-shirts: Chu-Ren Huang, the local arrangements chair, had assembled a team of 80 helpers, instantly identifiable by their brightly-coloured clothing. Their pervasiveness meant that any questions or problems were dealt with promptly and eagerly; but I suspect they must have been incredibly proactive, because I had few questions and saw few problems, in what was one of the most smoothly organised events I've been to in a while. The attention to detail was wonderful: from the large number of internet access points that meant you never had to wait to read mail, through the provision of different Taiwanese snacks and delicacies at each break and the reusable mug that everyone received as part of their conference pack, to the very thoughtful idea of

collecting name badges for recycling at the end of the conference.

The conference banquet was an exquisite ten course Chinese meal, accompanied by a Chinese wine that my tablemates seemed not to appreciate; but that was fine by me, since it made it much easier to acquire the taste. I suspect other participants enjoyed the wine too: I can't imagine what else might have provoked Jack Halpern, with no prior arrangement, to get up on the stage after Martin Kay's short presentation, to provide the assembled diners with a ten minute demonstration of unicycling. I suppose it's the logical next step if you want to cram even more vehicles onto the road, although I think it will take a bit more work before Taiwan's motor scooter manufacturers feel threatened.

The conference dosing session included what is slated to become a standard ceremony – the handing over of a 'Welcome to COLING' banner by Chu-Ren Huang to Maghi King, in charge of local arrangements for COLING 2004 in Geneva. See you there; I doubt this is the notion of badge recycling the organisers had in mind, but if there's a box of them shipped over to Geneva, mine is the one with several motor scooter tyre tracks across it, truly recycled.

FOR INFORMATION

Robert Dale is Professor in the Department of Computing in the Division of Information and Communication Sciences at Macquarie University, Sydney, Australia, and Director of the Centre for Language Technology.

Email: rdale@ics.mq.edu.au

Web: <http://www.mri.mq.edu.au/~rdale/>



Charles Fillmore with some orange T-

e

Workable Efficient Language Documentation: a Report and a Vision

Feature

Dafydd Gibbon, Universität Bielefeld



Dafydd Gibbon

Dafydd Gibbon reflects on experiences of working on projects documenting endangered languages in Africa – some of the lessons learnt and some hopes for the future.

software for use in adverse environments, e.g., in tropical rural and forest villages (see Figures 1 and 2).

Sample documentation of the endangered Ega language (from the Ivory Coast) is given at <http://www.spectrum.uni-bielefeld.de/LangDoc/EGA/>.

In order to disseminate further information about projects and ongoing work in the field, and to provide a showcase of best practice in endangered language documentation, the E-MELD portal has been established by the Linguist List (see <http://saussure.linguistlist.org/cfdocs/emeld/>).

Criteria for language documentation

But why 'language documentation' and, in particular, 'efficient' and 'workable' language documentation?

The goal of documenting all the world's languages is

just as important as documenting diversity in the biosphere or in geology: independently of the cultures associated with languages, each of the world's 7000 currently catalogued languages is a community-created abstract work of art in itself; it may be compared to a complex crystal, in that subtle fractal variations in structure distinguish it from all other languages and offer insights into unknown areas of the human cognitive potential. In the linguistic fieldwork community a very simple and practically motivated heuristic distinction is

now made between documentation of a language and its description. Briefly, the former constitutes the empirical foundations (recordings, 'sketch grammar'

Documentation projects: the WELD paradigm

In a number of projects since around 1997, one

funded by the Deutscher Akademischer Austauschdienst, another by the Deutsche Forschungsgemeinschaft and one by the Volkswagenstiftung, the Computation and Spoken Language group at Universität Bielefeld, Germany, has been developing efficient techniques for language documentation, and applying these to West African languages in pursuit of the 'Workable Efficient Language Documentation'

(WELD) paradigm. The techniques range from portable DAT recorder and laryngograph, and metadata logging with a Palm Pilot database application



Figure 1 shows a typical fieldwork situation in the Ivory Coast using portable DAT recorder and laryngograph, and metadata logging with a Palm Pilot database application

**Autumn
2002**

elsnet





Figure 2 shows an annotated video recording of a traditional African story-teller, made with open source (GPL) signal annotation software TASX developed by Jan-Torsten Milde at Universität Bielefeld (see <<http://tasxforce.lili.uni-bielefeld.de/>>).

with basic phonemic, morphological and grammatical analyses, and a basic lexicon) for a linguistic description, while the latter follows the patterns of scientific theory formation. From a more sophisticated perspective, empirical foundations of this kind necessarily interact in complex ways with theoretical assumptions, but there are urgent reasons for making compromises here and sticking to a simple distinction in the interests of efficiency and workability.

The urgency of language documentation

One of the urgent reasons for efficiency is that fieldwork on undescribed languages spoken, for instance, in rural or forest tropical areas or other relatively hostile climates requires complex logistics before the actual work can even start, therefore time is of the essence, and a high quality (thus time-consuming) linguistic description cannot always be made on location. Another consideration is that endangered languages are not likely to be with us for very long, and of the 7000 currently catalogued languages of the world hundreds (several thousand in the medium and long term) fall into this category (see <<http://www.ethnologue.org>>). High quality and efficient documentation is crucial for endangered

languages, because when a language is dead there is evidently no hope of collecting additional data and documentation must serve the purposes of science and the descendants of the extinct community in the permanent absence of native speakers.

It is evident to anyone with experience in language engineering projects that the size of the efficient documentation task is well beyond the abilities of individuals, projects, single consortia or research institutes. A vision needs to be developed for involving wealthy language engineering and computational linguistic communities and for spreading the idea of Workable Efficient Language Documentation beyond these communities to poorly equipped local scientific communities around the world with old computers, unstable electricity supplies, extremely expensive internet links (if any), and little if any contact with recent developments in the language and speech communities. Communities like these need tools which are workable in the local environment (not the latest heavy GUI software with proprietary applications and massive hardware requirements). But it is clear that the benefits of the WELD paradigm would not be one-sided – research and development on portability for such tasks would benefit many local language

e



communities around the world and have spin-off effects for portable speech and text technologies in other applications.

Towards a WELD charter

A Charter for the WELD paradigm would include at least the following five benchmark principles of comprehensiveness, efficiency, state of the art, affordability and fairness:

1. Language documentation must be comprehensive. In principle this means that language documentation must apply to all languages. But economy is a component of efficiency, and priorities must be set which may be hard to justify in social or political terms: if a language is more similar to a well-documented language than another language is, then the priority must be with the second.

2. Language documentation must be efficient. Simple, workable, efficient and inexpensive enabling technologies must be developed, and new applications for existing technologies created, which will empower local academic communities to multiply the human resources available for the task. A model of this kind of development is provided by the Simputer ('Simple Computer') handheld Community Digital Assistant (CDA) enterprise of the 'Bangalore Seven' in India (see <<http://www.simputer.org/>>), which could easily be incorporated into European and US project funding.

3. Language documentation must be state-of-the-art. In addition to using modern exchange formats and compatibility enhancing archiving technologies such as XML and schema languages, efficient language documentation requires the deployment of state of the art techniques from computational linguistics, human language technologies and artificial intelligence, for instance by the use of machine learning techniques for lexicon construction and grammar induction. The SIL organisation, for example, has a long history of application of advanced computational linguistic methodologies (see <www.sil.org>), and more research is needed here.

4. Language documentation must be affordable. In order to achieve a multiplier effect, and at the same time benefit education, research and development world-wide, local conditions must be taken into account. Traditional colonial policies of presenting 'white elephants' to local communities which must be expensively cared for and then rapidly become dysfunctional, must be replaced by inexpensive dissemination methods – at third world Internet prices, it can cost hundreds of euros to download a large, modern software package (not counting landline interruptions), and net-based registration and support

is unthinkable costly, as is wireless data transfer.

5. Language documentation must be fair. If a language community shares its most valuable commodity, its language, with the rest of the world, then the human language engineering and computational linguistic communities must do likewise, and provide open source software (also to reap the other well-known potential benefits of open source software such as transparency and reliability). The Simputer Public Licence for hardware and the Gnu Public Licence for software are useful references. The development and deployment of proprietary software (and hardware for that matter) and closed websites in this topic domain is a form of exploitation which is ethically comparable to other forms of one-way exploitation in biology and geology, for example in medical ethnobotany and oil prospecting.

Outlook

Naturally, things are not so simple in real life. Some of the principles in the WELD Charter outline may well be in conflict in some situations, requiring careful cost-benefit analysis. And there are in fact communities, fortunately not too common, who would be shocked at the thought of sharing their language with outsiders, just as there are of course R&D communities, unfortunately far more common, who would be shocked at the thought of sharing their resources with outsiders even in a context such as WELD. Intellectual property rights must be taken very seriously, of course, and the issues are far from simple. But the good news is that the dominance of these attitudes is slowly being replaced by WELD-friendly Open Archive, Open Resource, Open Source, and Open Data initiatives (just check the web here!), and that these are gradually being taken up by funding agencies as hallmarks of quality.

FOR INFORMATION

Dafydd Gibbon is Professor of Linguistics at the University of Bielefeld.

Email: gibbon@spectrum.uni-bielefeld.de

Web: <http://coral.lili.uni-bielefeld.de/~gibbon>

More information about the Language documentation projects can be found at <http://www.spectrum.uni-bielefeld.de/LangDoc/EGA>

Editor's note:

Readers interested in endangered languages and their documentation may also be interested in the announcement on page 11: 'A bold initiative to help preserve endangered languages'.

**Autumn
2002**

elsnet
•••••

Text Mining: a major advancement for Knowledge Management

The emerging commercially-based technologies

Jon Herring, University of Oxford



Jon Herring

On Wednesday 4 September, UK Euromap Language Technology held a seminar at the British Library to present emerging text mining technologies to the business community.

Euromap is an EU funded project dedicated to accelerating the transfer of human language technology research and development from the academic to the commercial sphere. In the UK, Euromap is based in the ITRI (Information Technology Research Institute) at the University of Brighton. Amy Neale of Euromap UK says: "One of the key aims of this seminar was to make business leaders aware of the potential solutions offered by new text mining techniques. These emerging applications are geared towards enabling companies to extract and make use of the huge amounts of locked-in, unstructured information in their organisations.

"Traditional data mining techniques are being married with developments in information retrieval and extraction," she added. "This allows the processing of textual information – after all four-fifths of the data on intranets and on the web is text not figures".

Six invited speakers, drawn from commercial research organisations, established research departments of UK universities, new spin out technology companies and the DTI, spoke to the

attendees about the new possibilities growing from ongoing research.

Jonathan Sage, leader of the eGovernment cluster at PwC consulting in Brussels gave an overview of the current state of play in Knowledge Management for large corporations, and his vision of how text mining techniques could be deployed. This was coupled with a demonstration of how machine translation technology had been taken up by PwC, and how the success of the project had convinced the organisation of the benefits of commercialising Human Language Technology (HLT) research.

Rob Gaizauskas, of the University of Sheffield's NLP group, provided a view from the forefront of IE research, outlining the difficulties and obstacles to building accurate and useful text mining applications, and surveying the different linguistic and computational techniques that are being brought to bear on the research questions. The talk included a demo of the TRESTLE system developed by Sheffield for GlaxoSmithKline, which mines a pharmaceutical news corpus to provide targeted summaries of important developments in drug patent approval.



Tom Khabaza, Programme Manager of the Advanced Data Mining Group at SPSS, demonstrated the successful integration of their Clementine data mining system with information extraction engines developed by the University of Brighton and LexiQuest (now incorporated into SPSS) and presented a range of potential applications of this kind of technology, from customer relation management and environmental conservation to fraud detection and crime analysis.

David Milward of Linguamatics in Cambridge



spoke about interactive information extraction – bringing the power of IE techniques to user-friendly search-engine-type queries. Daniel Brown of Applied Psychology Research showed how rule-based and probabilistic classification techniques could be combined to improve information management, and illustrated this with a case study of the data classification needs of one of APR's clients in the UK.

To round off the day, Margaret Dennis from the Department of Trade and Industry presented the opportunities offered by the EU's new 6th Framework programme (whose first call will appear in December this year) for collaborative IT and HLT research programmes.

The seminar not only provided an accessible and clear introduction to text mining and its possibilities for non-specialists, but was also a

good opportunity for building bridges between research and business, and a chance to identify possible future collaborations.

Conference report
(cont.)

FOR INFORMATION

Jon Herring is Lecturer in Linguistics at the University of Oxford and a part-time PhD student at the ITRI, University of Brighton

Email: Jon.Herring@itri.bton.ac.uk

Web: <http://www.itri.bton.ac.uk/~Jon.Herring>

More information about Euromap can be found at <http://www.itri.bton.ac.uk/projects/euromap> or <http://www.hltcentral.org/euromap>.

Email: euromap@itri.bton.ac.uk

EACL to include new paper submission category

Announcement

The next EACL meeting, to be held in Budapest, Hungary in April next year is going to pilot a new approach to paper submissions, allowing it is hoped, the conference to have a wider scope than the traditional EACL and ACL diet of mainstream work. This year's conference includes two separate categories of paper submissions, with the usual paper category for papers up to eight pages describing completed work of an original nature. In addition, however, this year there is to be a category of 'Research notes and demos' (adding to the previously existing "demo" category), to include work in progress, project status reports, unevaluated results or system summaries (with or without demos). This latter category will be up to four pages.

It is hoped that this will revive interest in submitting papers to EACL, which has developed something of a reputation as a very difficult place to get papers accepted. It should also allow researchers working in less well established subfields and countries with a shorter history of research within the field to air their work before a wider audience than they might otherwise be able to do, without such pressure to conform to the norms within the field. The inclusion of 'project notes' in conferences such as COLING in recent years has been a great success, with the resulting programmes enormously varied while still of a generally very high standard. Although it may be easier to get papers accepted, and the requirements in

terms of the level of originality and status of completion may be lower, it is anticipated that the quality, both of the work reported and the presentation of this work, will remain high.

Deadlines for submission of papers will also be staggered, with the deadline for full paper submission the 15th November (registration by 10th November) and the deadline for research notes and demos three weeks later, on 6th December (registration by 1st December). Full details of the EACL submission requirements can be found on the conference web site (see below).

FOR INFORMATION

EACL 2003

Dates: April 12th-17th 2003,

Location: Budapest, Hungary

Web: www.conferences.hu/EACL03

Programme Chairs: Ann Copestake, Jan Hajic

Email: aac@cl.cam.ac.uk, hajic@ufl.mff.cuni.cz

Local arrangements: Ferenc Kiefer

Email: kiefer@nytud.hu

**Autumn
2002**

elsnet
.....

SmartKom – Intuitive Multimodal Human-Machine Communication

Wolfgang Minkler and Peter Regel-Brietzmann
DaimlerChrysler AG, Research and Technology, Germany

There is a growing need for more intuitive and efficient interfaces to computing devices — especially for users with limited technical experience or physical disabilities and in mobile or time-critical interaction.

SmartKom (www.smartkom.org) is a long-term research effort funded by the German Federal Ministry for Education and Research (BMBF) in its funding area Human Technology Interaction (www.dlr.de/IT/IV/MTI). Started in 1999, the project aims at developing advanced concepts for intuitive human-centered computer interfaces. The lead contractor of the project consortium is the German Research Center for Artificial Intelligence (DFKI). Partners from industry include DaimlerChrysler Research and Technology, Philips, Siemens and Sony.

SmartKom aims to exploit one of the major characteristics of human-human interactions: the co-ordinated use of different code systems such as language, gesture, and facial expressions for interaction in complex environments. A mixed-initiative approach allows intuitive access to knowledge-rich services. Three user interface paradigms – spoken dialogue, graphical user interfaces, and gestured interaction – are merged to achieve truly multimodal communication. Natural language interaction in SmartKom is based on speaker-independent speech understanding technology. For the graphical user interface and the gestured interaction SmartKom communication assistants do not use a traditional WIMP (windows, icons, menus, pointer) interface; instead, they support the natural use of gestures.

SmartKom's interaction style breaks radically with the traditional desktop metaphor. The multimodal system is based on the situated delegation-oriented dialogue paradigm (SDDP): the user delegates a task to a virtual communication assistant, visible on the graphical display. With more complex tasks, this delegation cannot be achieved in a simple command-and-control style. In a collaborative dialogue between the user and the system – which is represented by a lifelike character – specifications of the delegated task and of possible plans are worked out. In contrast to task-oriented dialogues, in which the user carries out a task with the help of the

system, with SDDP the user delegates a task to an agent and helps the agent, where necessary, in the execution of the task.

The abilities of the SmartKom assistants are currently being tested within three real application scenarios. Thus, the project examines the flexible integration of functionalities and hardware into a homogeneous system.

SmartKom Public is a multimodal communication booth that offers a wide range of advanced communication devices such as a document camera for sending images of real world objects. In the near future, this type of booth may provide information about train and flight schedules, city maps, hotels and cinemas at airports, train stations and other public places. Access to SmartKom Public may be granted upon authentication and payment by credit card.



SmartKom Public – a design study

SmartKom Home/Office is a personal domestic computer workstation that is not limited to address management, document processing and information access. SmartKom Home/Office assistants constitute a real support in everyday life, since they offer an intuitive user interface to a wide range of home entertainment and other domestic devices (e.g. radio, television and video-recorder). The systems are accessed through a portable webpad. The user operates SmartKom Home/Office either in lean-forward mode, with co-ordinated speech and gestured interaction, or in lean-back mode, with voice input alone.

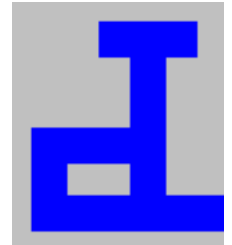
FOR INFORMATION

Wolfgang Minkler and **Peter Regel-Brietzmann** are members of the scientific staff at DaimlerChrysler, Research and Technology (Germany).

Email: wolfgang.minkler@daimlerchrysler.com
peter.regel-brietzmann@daimlerchrysler.com

Web: <http://www.smartkom.org>

SIGdial Website: <http://www.sigdial.org/>



EDILOG: dialogue and dancing

Conference report

Paul Piwek, ITRI, University of Brighton

[Setting: Researcher Mary meets her colleague John at Edinburgh Waverley railway station. Mary has arrived by train, and John is about to catch a train. It's a sunny day, early September.]

Mary: Hi John, what has brought you to Edinburgh?

John: Hello Mary. I've just been to the EDILOG workshop.

Mary: The EDI what?

John: EDILOG, it's the sixth one in a series on the semantics and pragmatics of dialogue. The first one was held in Munich and called Mundial. That started a tradition of adding "dial" and later "log" as an affix to part of the host city's name.

Mary: Sure, but what was this workshop all about?

John: Well, it began in 1997 as a gathering of a small group of mostly formal semanticists interested in modelling dialogue. But since then it has grown into a forum for researchers from a wide variety of fields who are studying dialogue.

Mary: Can you be a bit more specific?

John: Well, actually this year the range of perspectives in the 24 regular talks and eight posters and demonstrations was nicely reflected in the presentations by the four invited speakers. There was Susan Brennan presenting results from empirical studies into whether speakers and addressees adapt to one another in human-human dialogues. Stanley Peters delved into the foundations of modelling dialogue and linked this to ongoing work on building dialogue systems. Furthermore, Manfred Pinkal discussed how dialogue might help to bridge the gap between applications and formal/theoretical work in semantics and pragmatics. Although he pointed out that there is no easy road to success, he gave examples of some projects where creative thinking helped to forge links. Finally, Enric Vallduví argued that dialogue is the place where context change in communicative interaction and information structure (theme-rheme articulation) meet. He



A dialogue at Arthur's Seat

illustrated this by discussing some entertaining dialogue fragments

Mary: All right, but I bet that besides all this food for thought, there were also some less taxing activities?

John: I must confess it wasn't all hard work. The organizers didn't forget to include some lighter moments in the programme. From a murder mystery tour through the city, a traditional Ceilidh at the workshop party to a healthy climb up Arthur's Seat overlooking Edinburgh or, alternatively, a visit to the pub.

Mary: I think your train is almost leaving; but tell me first what will next year's workshop be called?

John: Well, have a guess; but I'll tell you this, next year it will take place in Saarbrücken.

[... John manages to slip into the train which departs almost immediately after he closes the door behind him]

FOR INFORMATION

Paul Piwek is Research Fellow at the ITRI, University of Brighton

Email: paul.piwek@itri.bton.ac.uk

Web: <http://www.itri.bton.ac.uk/~Paul.Piwek/>

**Autumn
2002**

elsnet



The UNISYN lexicon for speech synthesis

Compiled by *Lynne Cahill* with input from *Sue Fitt* and *Steve Isard*, CSTR, University of Edinburgh

When synthesised speech started coming into general use, the most frequent complaint was that the voices sounded too robotic, not human enough. The field has now advanced to such a level that the complaints are now much more subtle, such as “Why do they all have American accents?” That such a stage has been reached should be enormously encouraging to workers in the field, and even more encouraging is the fact that many speech synthesis systems now address these concerns, providing a range of different accents as well as different languages. One example of this is the Festival speech synthesis system, developed at CSTR, Edinburgh. Festival already comes with a choice of different ‘voices’, allowing users to choose from a range of female or male, UK or US varieties of English, as well as other languages such as Spanish and Welsh, by means of separately defined lexicon and realisation modules. It offers a full text-to-speech system with various APIs, as well as an environment for development and research of speech synthesis techniques. It is written in C++ with a Scheme-based command interpreter for general control.

Until now, the Festival synthesiser, like most others, needed a separate lexicon for each accent, dialect or language it speaks. Large lexicons are only available for General American and Southern British (RP) English, greatly limiting the scope of current synthesis. Creating a large lexicon for every regional accent of English would be costly in both time and effort. The intention behind UNISYN is to render this cost unnecessary by exploiting regular connections between the pronunciation patterns of different accents to permit the definition of more than one accent within a single lexicon.

The UNISYN lexicon, developed by Sue Fitt and Steve Isard, does just this, making use of ‘keysymbols’, which operate as ‘metaphonemes’, generalising over the realisation of sounds in different accents of English. For example, the RP pronunciation of the vowel sound in ‘bath’ is the same as the RP vowel sound in ‘car’, and distinct from ‘trap’, while in Northern English ‘bath’ and ‘trap’ are the same as each other and distinct from ‘car’. To generalise across these accents, a keysymbol needs to be defined for the vowel sound

in ‘bath’, which is realised as the phoneme /ɔ:/ in Northern English and /ɑ:/ in RP. A separate symbol is used for ‘car’, and is realised onto /ɑ:/ in both RP and Northern English, and there is a third symbol for ‘trap’, realised as /ɒ/ in both accents. (The above phonological representations are all given in the SAMPA machine readable phonetic alphabet.)

It should be noted that the distinctions are between accents, not dialects, as it is only the pronunciation that can differ within the UNISYN lexicon. The Festival system allows users to add their own ‘voices’ by providing their own data files, which can be defined at a number of different levels. For example, while one user may wish to define a whole set of lexicon and pronunciation files, another may just provide a single pronunciation file to deliver a variation in the pronunciation of a certain set of words. This work is being continued by Alan Black at Carnegie Mellon University, under the Festvox project.



Sue Fitt – UNISYN developer

The Unisyn lexicon is a master lexicon transcribed in keysymbols, a kind of metaphoneme, which allows the encoding of multiple accents of English. The lexicon is accompanied by a number of Perl scripts, which transform the base lexicon, via phonological and allophonic rules and other symbol changes, to produce output transcriptions in different accents. The rules can be applied to the whole lexicon, to produce an accent-specific lexicon, or to running text. Output can be displayed in keysymbols, SAMPA or IPA.

Recent work by Cahill and Tiberius (<http://www.itri.bton.ac.uk/projects/metaphon>) has considered defining metaphonemes (essentially the same as the keysymbols in UNISYN) →

A large, white, stylized lowercase letter 'e' is centered on a dark blue square background. The 'e' is a simple, rounded font style.

across closely related languages – Dutch, English and German, primarily. Even though this cross-lingual work is intended to form part of a broader lexicon system, including more distinctions than just pronunciation, nevertheless the respective moves in this direction offer the hope that advances in speech synthesis will become much more easily transferred both across accents and across different (but related) languages.

ACKNOWLEDGEMENTS

UNISYN was sponsored by EPSRC
Festival was sponsored by EPSRC, Sun Microsystems, AT&T Labs – Research, and BT.
METAPHON was sponsored by ESRC

FOR INFORMATION

For further details of the UNISYN accent-independent lexicon (including free download):

Web: www.cstred.ac.uk/projects/unisyn

Email: sue@cstred.ac.uk

For more details about the Festival speech synthesis system (including free download):

Web: www.cstred.ac.uk/projects/festival

Email: festival@cstr.ed.ac.uk

For more details of the Festvox project:

Web: www.festvox.org

Email: awb@cscmu.edu

Feature (contd..)

A bold initiative to help preserve endangered languages

Press release

To help explore and record linguistic diversity across the globe, a British foundation has provided £20,000,000 over ten years to create an international scholarly program to study endangered languages.

The scale of the funding is commensurate with the urgent – and enormous – threat to the world's linguistic diversity. Many of the languages that will be studied are linguistic isolates. All are very nearly extinct. They have never been adequately analysed or recorded, and they are typically spoken only by a few elderly people. These languages – and their speakers – deserve to be remembered, and to take their place in history. At the same time, this worldwide project to preserve crucial knowledge about the world's linguistic heritage will vitally illuminate the history of how humanity settled the earth.

The Lisbet Rausing Charitable Fund aims to support research in the humanities and the social sciences. This grant, together with other family benefactions amounting to many millions of pounds, is intended by the Hans Rausing family to help British universities maintain the highest standards of academic scholarship.

When deciding to secure the participation of SOAS in this programme – a process that took many months of consultation – the Fund's trustees expressed the greatest confidence in the achievements and potential of the School, and in the enthusiasm and dedication of its scholars and leaders. The trustees were impressed by the

fit between their own profound concern at the threat to knowledge of linguistic and cultural diversity globally, and SOAS's long-standing and distinguished study of small languages in Africa, Asia, the Middle East, and elsewhere. The Fund's trustees also share with SOAS a commitment to the highest ethical standards when co-operating with small language communities – people who are often marginalized and dispossessed.

Part of the grant will underwrite an academic programme within SOAS, utilising SOAS's staff and facilities. It will train field-workers and deepen knowledge of endangered languages through specially designed courses in field linguistics generally and endangered languages in particular as well as by co-ordinating scholarly activity, publicity and consultation in the field. But the bulk of the fund will be administered by SOAS to provide grants to scholars throughout the world to document and analyse endangered languages.

Professor Colin Bundy, Director and Principal of SOAS, voiced unqualified delight at the news of the award. "SOAS was founded in 1916 as a specialist institution for the study of languages in Asia, and later in Africa. We created the first British linguistics department (in 1932) and our Library was identified in 1961 as a national resource for the study of Africa and Asia. Our history, mission and ethos equip us for this visionary project." He stressed that in addition to the School's regionally defined departments concentrating on language and culture its range of disciplinary

**Autumn
2002**

elsnet
.....

departments – such as anthropology, history, linguistics – offered a rich opportunity for becoming a world leader in the documentation and study of endangered languages.

SOAS and the Fund together will underwrite the infrastructure to manage this grants programme. This means that other families, foundations and companies that would like to donate to this cause will have the security of knowing that 100% of their money goes directly to the recording and study of nearly extinct languages. The costs of research and documentation to ensure that full knowledge of a language and its use are preserved will vary, but the average is about £150,000: readers may wish to donate money to this profoundly important cause – before those thousands of the world's languages (well over 50% of the total) that are now highly threatened, disappear forever. No sum is too small, and all money donated will go directly, fully, and only to the cause of recording near-extinct languages – and thus save a unique world heritage.

Applications will be invited from researchers – who might include suitably qualified research students or postdoctoral candidates, as well as senior and established academics – with qualifications in and, ideally, experience of field linguistics. The core of the programme will probably be grants to support more or less elaborate projects for the documentation of individual or closely related endangered languages, involving one or more researchers and receiving support for up to three or, in exceptional circumstances, four years. In the first instance applicants will be expected to submit a relatively brief

Summary Proposal Form. These will be assessed and those which appear to conform to the programme's expectations as to importance and quality will be invited to submit a more detailed application.

It is anticipated that in this first 'round' the date for submission of Summary Proposals will be mid-October 2002; invitations to submit detailed applications will be despatched in late November 2002; and the closing date for detailed applications will be early January 2003.

Detailed applications will have to conform to a variety of standards (including ethical and technical standards), which will be specified on the website.

FOR INFORMATION

Web: www.eldpsoas.ac.uk

Payments to SOAS can be made direct to the School's bank at:

National Westminster Bank plc
94 Moorgate
London EC2M 6XT
Sort Code: 56-00-23
Account No: 08622655

All general enquiries should be addressed to Mary O'Shea at SOAS

Tel: +44 (0)7898 4075

Email: mo2@soas.ac.uk

Announcement

HOARSE Opportunities

The HOARSE (Hearing, Organisation And Recognition of Speech in Europe) IHP (Improving Human Potential) NETWORK is a research training network funded by the EU fifth Framework 'Improving Human Potential' programme from 1st September 2002 for 4 years. HOARSE will fund both postgraduate (i.e., Ph.D. students) and postdoctoral researchers. The scientific objectives of HOARSE are to gain a better understanding of speech production and hearing and to use this understanding to explain the perceptual organisation of sound and improve speech technology. The project has five inter-related themes:

- How are sound mixtures perceptually organised, and how can this Auditory Scene Analysis be used in speech recognition?
- How does the Auditory System handle reverberant conditions, and how can models of this processing be used for speech enhancement?
- How is speech production related to speech perception and cerebral speech processing, and how can this knowledge be integrated into speech recognition systems?
- How can Automatic Speech Recognition algorithms take

advantage of the work in Themes 1 and 2, for use in natural listening conditions?

- How can the results of other themes be exploited in speech recognition applications which require robust performance in adverse conditions and/or processing of sound mixtures?

HOARSE research is multidisciplinary and a background in Acoustics, Audition, Signal Processing, Phonetics, Linguistics, Computer Science or Mathematics may be suitable. The HOARSE partners are: Speech and Hearing Research Group, Department of Computer Science, University of Sheffield (coordinator); Ruhr Universität Bochum; Daimler-Chrysler Research, Ulm; Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology; FINIDIAP, Martigny; University of Keele; Wire Communication Lab, University of Patras.

FOR INFORMATION

Web: www.hoarsenet.org

Email: p.green@dcs.shef.ac.uk



Opinion

Opinion Column

Kenneth Church, AT&T Labs Research, USA

My last column suggested we discuss strategy at this summer's meetings. The idea that "more data is better data" has been floating around for a while but it seems to be picking up steam (again). There is a lot of excitement about the web, which is not only large, and growing, but also contains a rich structure of hypertext links. I am going to suggest here a bait and switch strategy, where the public Internet is the bait, but the real target is something larger and more valuable, but more elusive.

One of the two EMNLP best papers (there was a tie), Keller *et al*, 'Using the Web to Overcome Data Sparseness', showed that probability estimates obtained from larger corpora (e.g., Google) were better than those obtained from smaller corpora (e.g., British National Corpus) for predicting psycholinguistic judgements. It was also suggested in the conclusions that web counts might be better than standard smoothing techniques such as back-off that are often used for language modeling. I think this is really exciting. It looks as though performance on a broad range of computational linguistics tasks will improve as we collect more and more data. The rising tide of data will lift all boats.

Another EMNLP-02 paper, Brill *et al*, 'An Analysis of the AskMSR Question-Answering System', showed that one can do remarkably well in TREC question answering competitions by using a search engine like Google and very little else. Norvig not only made more or less the same point during his ACL-02 invited talk, but he also highlighted Google's ability to find interesting sets of words. Given 'cat' and 'dog', labs1.google.com/sets returns a list of animals. Given a few countries, the page returns more countries, often from similar parts of the world. We used to try to do similar things a decade ago, but the results were not nearly as good, probably because we were working with relatively tiny corpora in the sub-billion-word range.

Of course, the web is hardly a balanced corpus. Norvig made this point rather dramatically with the seed words, 'cat' and 'more', which caused <http://labs1.google.com/sets> to return a list of Unix commands! Unix commands and many other subjects, especially taboo subjects, are over-represented on the web. There has been a lot of discussion in corpus linguistics on balance, but overall, the results mentioned above suggest that while balance is desirable, size is even more desirable. Of course, the trade-off depends on the application. I would have expected balance to be relatively

important for predicting psychological judgements and size to be relatively important for language modelling. It is surprising that size seems to be winning out over balance in both cases.

Large as the public Internet is, there are even larger opportunities. Changing copyright laws in various ways might unlock vast resources. Publishers like www.lexisnexis.com have impressive collections that may well surpass the public Internet. Private intranets and telephone networks have even larger sources of linguistic data. Of course, most of the data on private Intranets cannot be distributed outside the Intranet, and most of the telephone traffic cannot even be recorded. But attitudes are changing. Voice-mail took a while to catch on. It used to be considered rude to have an answering machine; now, it is considered rude not to have one. It is hard to know how much speech could be recorded, but between answering machines and call centres (<http://www.informationweek.com/story/IWK20020808S0010>), perhaps 10% can be recorded.

How large is large? According to Keller *et al*, Google is nearly 1000 times larger than the British National Corpus (BNC); that is, Google has about 100 billion words compared to the BNC's 100 million. Telephone networks, of course, are much larger. According to http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/trends.html, there are roughly 200 million telephone lines in America, one for each person in the country. Each line is used about an hour a day. If we assume that a second of speech corresponds to roughly a word, the American telephone network generates about 10 Google collections per day. It is hard to say how large the private intranets are, but right now, at least, revenue-wise, the data networks are roughly comparable to the voice networks, though data is growing faster than voice.

In the past, recording all this data would have been prohibitively expensive. But thanks to Moore's Law, storage costs have been falling faster than transport for some time and will continue to do so. Even at current prices, if I am willing to pay for a long-distance telephone call (5 cents/minute), I might as well pay for the disk space to keep the speech online (0.5 cents/minute). Similar comments hold for web pages. Why flush a page if there is any chance that it might be requested again? Over time, web caches will look more like web crawlers. Go find the pages that I might ask for and keep them for ever.



**Autumn
2002**

elsnet



Opinion Column
(contd.)

The proposed bait and switch strategy is to use the public internet as the bait to develop and test and socialise new ways of extracting value from large linguistic repositories. The value to society, though, is when these solutions are applied to the private repositories that we care about. (No one cares about data that everyone can have, just as Groucho Marx doesn't want to be in a club that would have him.)

To a large extent this strategy encourages the research community to keep doing more of all the great stuff that we have been doing. There will be more interest in papers that not only report performance on currently available corpora, but also report how well the techniques port from one corpus to another. There will also be more interest in papers that report how well performance scales with corpus size. Of course, size doesn't always help, and all the data in the world will not solve all the world's problems. It would be useful to know when more data will help and when it is better to do something else (e.g., a revival of linguistic representation). There have been many examples of these kinds of papers in the past, and hopefully there will be many more.

As for investments in infrastructure, in addition to

traditional data collection efforts that are focused on public repositories, we ought to think about private repositories as well. Most of us, for example do not keep voice mail for very long, though I have been using Scanmail (Hirshberg *et al*, Eurospeech-2001) to copy my voice mail to my email, and like many people, I keep a lot of email online for a long time. Unfortunately, the tools for searching email archives and other private repositories are not as good as the tools for searching public repositories. We could make a huge difference in the size of private repositories by making it more convenient to capture private data, and by demonstrating that there is value in doing so.

FOR INFORMATION

Ken Church is Head of the Department of Data Mining at AT&T Labs Research in New Jersey

Email: kwc@research.att.com

Web: <http://www.research.att.com/~kwc>

Call for book
proposals

Elsevier Science – Call for Proposals in Linguistics and Computation

Elsevier Science is looking to expand its distinguished linguistics programme into the field of Linguistics and Computation. We are interested in a range of proposals including journals, books and book series.

Why publish with Elsevier Science?

- A widely read and respected forum for your work
- Renowned standards of excellence in publishing: *Journal of Pragmatics*, *Lingua*, *Computer Speech and Language*, *Handbook of Logic and Language* and the *NHLS* and *CRiSPI* book series
- Articles available electronically via ScienceDirect – the world's best internet resource for scientific, technical and medical information
- Generous discounts on Elsevier Science publications for contributors
- World-Wide Marketing and Exhibitions

For more information on publishing with Elsevier Science visit www.authors.elsevier.com or contact:

Sarah Oates, Publishing Editor for Linguistics,
Elsevier Science,
Langford Lane,
Kidlington, Oxon OX5 1GB
United Kingdom

Tel: +44 (0) 1865 843562

Fax: +44 (0) 1865 843951

Email: SOates@elsevier.com

Web: <http://www.socscinet.com/linguistics>



Future Events

- Oct 12** *International Conference on Multimodal Interfaces (ICMI 2002)*: Pittsburgh, USA
Email: celine@cs.cmu.edu URL: www.is.cscmu.edu/icmi
- Oct 23-24** *European Commission's Sixth Framework Programme Information Days*: Luxembourg
Email: adriana.bini@cec.eu.int URL: www.cordis.lu/ist/ka3/news.htm
- Nov 4-7** *LISA Forum Europe: Standards in Localization and Translation*: Heidelberg, Germany
Email: lisa@lisa.org URL: www.lisa.org/events/2002europe
- Nov 6** *TRAIN-IT Workshop at IST 2002*: Copenhagen, Denmark
Email: bechmann@izet.de URL: 2002.istevent.cec.eu.int
- Nov 12** *IBERAMIA Workshop: Multilingual Information Access and Natural Language Processing*: Seville, Spain
Email: anselmo@lsi.uned.es URL: sensei.lsi.uned.es/iberamia-mlia/
- Nov 14-15** *6th EAMT Workshop: Teaching Machine Translation*: Manchester, UK
Email: Harold.Somers@umist.ac.uk URL: www.ccl.umist.ac.uk/events/eamt-bcs
- Nov 16** *CSCW2002 Workshop on Storytelling and Collaborative Activities*: New Orleans, USA
Email: leonie.shaefer@fit.fraunhofer.de URL: www.acm.org/cscw2002/prog-workshops.html#w2
- Nov 21-22** *24th Translating and the Computer Conference*: London, UK
Email: barbara.hobbs@aslib.com URL: www.aslib.com/conferences
- Nov 27-29** *ISCA Workshop "Advanced ASR for Telecom Applications"*: Avignon, France
Email: Els.denOs@mpi.nl URL: lands.let.kun.nl/SMADA
- Nov 29** *Computational Linguistics in the Netherlands (CLIN 2002)*: Groningen, The Netherlands
Email: clin@let.rug.nl URL: odut.let.rug.nl/c lin2002
- Dec 9** *The 2002 IEEE International Conference on Data Mining*, Maebashi City, Japan
Email: otani@maebashi-it.ac.jp URL: kis.maebashi-it.ac.jp/icdm02
- Jan 15-17** *Fifth International Workshop on Computational Semantics*: Tilburg, The Netherlands
Email: computational.semantics@kub.nl URL: let.kub.nl/research/TI/sigsem/iwcs/iw cs5

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

A message from the new editor

As this is my first attempt at desktop publishing I hope you will bear with me and forgive me any glaring errors or omissions. I have tried to include an interesting balance of articles, announcements etc., but the future success of any newsletter like this requires at least some input from the readers. I would like to encourage everyone who has an interest in the field to send me letters, articles and announcements of all kinds, whether about your own research project, something you've seen that interested you or a new course or book that you would like to tell the community about. Even if you don't have the time (or inclination) to write an article yourself, I would still be grateful for suggestions or nominations of likely victims.

I would like to thank my predecessor, Jenny Norris, for leaving things in a sensible state for me to take over, and for her help with the handing over process. I would also like to thank the ELSNET organisers in Utrecht and Geoffrey Sampson at Sussex who have made me feel welcome in my new job.

Autumn
2002

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,
LIMSI-CNRS (F)

Executive Board

Steven Krauwer,
Utrecht University (NL)
Niels Ole Bernsen, NIS,
Odense University (DK)
Jean-Pierre Chanod,
XEROX (F)
Björn Granström,
Royal Institute of
Technology (S)
Nikos Fakotakis,
University of Patras (EL)
Ulrich Heid,
Stuttgart University (D)
Denis Johnston, BT
Adastral Park (UK)
Joseph Mariani,
LIMSI-CNRS (F)
José M. Pardo,
Polytechnic University of
Madrid (E)
Tony Rose, Reuters (UK)
Geoffrey Sampson,
University of Sussex (UK)
Antonio Zampolli,
University of Pisa (I)

ELSNET P 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 Erlang-en-Nürnberg -FORWISS
D Ruhr-Universitaet 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 LIMSI-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 DaimlerChryslerAG
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 Compuer
NL Knowledge Concepts BV
NL Sopheon NV
NL IP Globalnet Nederland BV
RU ANALIT Ltd
RU Russicon Company
S Sema Infodata
S Telia Pionator AB
UK Imagination Technologies plc
UK Canon Research Centre Europe Ltd
UK Sharp Laboratories of Europe Ltd
UK BT Adastral 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 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

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

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