elsnews

The Newsletter of the European Network in Human Language Technologies

Summer 2003

EACL 2003: The Hungarian Experience

António Ribeiro European Comission Joint Research Centre, Ispra, Italy

This year EACL joined the European trend towards the enlargement of the European Union by organising its first conference in a Eastern European country.

Hungary was in deed a memorable first stop for the celebration of the tenth EACL conference. I must say I was rather surprised by the high quality of the talks overall in this particular conference, even after having attended EACL and ACL conferences for several years now. Also, I think every participant was impressed with the host city: Budapest is a remarkable classical city going through a transitional period, with old and new living side by side. With the bustling Moscow Square transport hub and the bohemian Franz Liszt Square café esplanades; with its 100 year old underground to the City Park, and the Danube 'waltzing' through both sides of the city; with the castle on a hill on the Buda side, overlooking the classical buildings in Pest on the other bank of the Danube.



The Chain Bridge in Budapest links Buda and Pest over the Danube.

The city indeed has so much to offer (yes, I am including those special Hungarian thermal baths, which seem to be ubiquitous in the city, and which so many participants could not resist).

Back to the conference. EACL 2003 took place between Saturday 12th and Thursday 17th April. Around 400 people attended this conference, with about ten per cent of the participants coming from Eastern European countries. This proved to be a great occasion to get to know about the current research efforts and developments in these countries. The paper acceptance rate was about 26%, with 48 papers accepted out of 181 submitted, the highest number of submissions an EACL conference has had so far. The conference included one day of tutorials two days of workshops, and a three day main conference programme with three parallel sessions, which made me rush from

one session to another. Fortunately the organises allowed some time for this. With a set of twelve workshops to choose from, it was not surprising to see that suddenly the conference site started to get busier and more crowded by the time the workshops started.

This year, EACL also included a new session for Research Notes and Demos. This session gave the chance for many to show their demos or to present some work still in the early stages of development in order to get feedback. Some of the demo rooms were very busy with participants lining up or crowding around some

| Inside this issue: | |
|--|----|
| Let ther e be more sound quality Florian Hammer | 3 |
| LangTech2003 | 4 |
| SIGDIAL Ame Jönsson | 5 |
| Copus Linguistics 2003 Andrew Roberts | 6 |
| New ELSNET members | 8 |
| Opinion Annie Zaenen | 10 |
| ENABLER Network | 11 |
| ParGram: Developing parallel grammars Helge Dyvik | 12 |
| Calendar | 1: |
| | |

12.2





exhibitors to get to know and talk to the demo presenters. These sessions were organised as the 'third parallel session' of the main conference, so that, even if you were not very interested in one of the main conference papers, you could always pop into the demo sessions to visit some of the exhibitors.

As I pointed out earlier, the general quality of the papers was rather good and leaves me unable to give special mention to any particular paper, although there were some which I found more interesting

to my research interests.

The conference organisers offered us two interesting invited talks. The first invited talk, "Multilingual Access to Large Spoken Archives" by Doug Oard from the University of Maryland, USA, gave us a presentation on a project which is to provide multilingual and cross-language access to a large collection of interviews with survivors of the Holocaust. A quite impressive task ahead. Just before the conference closing, John Nerbonne, from the University of Groeningen in the Netherlands, gave a talk with an unusual insight into the "Linguistic Variation and Computation in the Netherlands", where he established a link between Dutch geography and the lexicophonetic distribution across the country.

The Student Research Workshop took place during the main conference. It has established itself as a particularly important conference session as it has allowed students to present their work, receive feedback from top researchers, and have the chance to meet other researchers. Indeed we all had the opportunity to attend some very lively and interactive presentations

With so much activity going around we all deserved a nice conference banquet. It was set in a unique environment, aboard the 'Europa' boat. It was a nice coincidence as Hungary was celebrating the "yes" result in the referendum on entry into the European Union. The boat cruised along the Danube with impressive views over Buda and Pest in the evening, and everyone had the chance to enjoy some nice Hungarian food and some dancing.

because they were closer EACL participants enjoying the view over Budapest. Left to right: Mary Wood, Donna Harman, Rob Gaizauskas, Mark Greenwood, and Jimmy Lin. Picture courtesy of Rob Gaizauskas.

> This year's EACL included a contest for the best paper award. Participants were invited by Jan Hajic, one of the programme co-chairs, to vote for the best paper. Ann Copestake, the other programme co-chair, revealed the winner in the closing session and the award was granted to Geert-Jan Kruijff, University of the Saarland, and Jason Balridge, University of Edinburgh, for their paper on "Multi-Modal Combinatorial Categorial Grammar".

All in all, I think all conference chairs did an excellent job by setting up another of these top quality EACL conferences. We all learnt that EACL is now going to be organised every three years, as ACL will come to Europe also every three years. Thus, there will probably be just one year in three without a major computational linguistics conference in Europe. Anyway, I am looking forward to next year's ACL which is coming to Europe and will be hosted in the lively Spanish city of Barcelona.

FOR INFORMATION

António Ribeiro is currently researcher at the group of Language Technologies, in the Institute for the Protection and Security of the Citizens, at the Joint Research Centre of the European Commission, Ispra, Italy.

Email: Antonio.Ribeiro@jrc.it

Web: www.jrc.cec.eu.it/langtedh/ar.html

EACL web site: acl.ldc.upenn.edu/eac12003



Editor: Lynne Cahill.

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

ISSN 1350-990X

© ELSNET 2003

FOR INFORMATION Contributions to ELSNews, and corrections, should be sent to lynneca@cogssusx.ac.uk Tel: +44 1273 678659 Fax: +44 1273 671320

Material for the next issue is due: 26th September 2003





Workshop report

Let There Be More Sound Quality

Impressions from the ISCA Tutorial and Research Workshop on "Auditory Quality of Systems", April 23-25, 2003.

Florian Hammer

Telecommunications Research Center, Vienna

The ISCA tutorial and research workshop on Auditory Quality of Systems was successfully organised by the team around Ute Jekosch and Sebastian Möller of the Institute of Communication Acoustics of the Ruhr-University Bochum (IKA), Germany. Fascinating architecture built at the site of an old coal-mine served as the environment for this multidisciplinary event. The Akademie Mont-Cenis provided the optimal configuration concerning the lecture hall size, space for informal information exchange, and comfortable accommodation.

The program was divided into sessions on topics like sound quality assessment, quality assessment and prediction in telecommunications, quality of voice-over-IP connections, methodologies and methods, quality of virtual and real environments, quality of music, quality of speech technology, and usability issues.

In the first keynote talk, Prof. Blauert (IKA) pointed out the need to distinguish the terms *sound/speech quality* and *sound/speech transmission quality*. He gave the following definition of speech quality which is based on the work of Jekosch:

Speech Quality is the result of an assessment of the adequacy of a speech sample – considering all of its recognised and nameable features and feature values – namely, as to which amount this speech sample camplies with a reference arising from aspects such as individual expectations and/or social demands and/or pragmatic

necessities – considering all recognised and nameable features and feature values of the reference.

After talks on sound quality assessment including expectation-based evaluation, attention was drawn to telecommunications applications and, as a particular research field, voice-over-IP. There was a plea for the use of corrupted speech data, an investigation of the importance of VoIP-packets, and an exploration of whether time-varying degradations are additive to stationary degradations in heterogeneous networks. The methodologies and methods session focussed on the judgement of sound quality from a psychological point of view (choice models, direct scaling) and paired-comparison tests of MP3-audio quality. Then, the quality of virtual and real environments was discussed, starting from the impact of spatial distribution of reflections on the auditory quality and character in virtual acoustic environments, to sensed presence in virtual environments, quality of head-related transfer functions (HRTFs), and a demonstration of a software platform for the evaluation of spatial attributes of reproduced and interactive 3D sound (cf. Virtools).

Based on a collection of topics related to auditory quality of systems, a fruitful discussion ensued. During this discussion, issues were raised that might be important for further research, e.g, deficiencies *vs* benefits, sensitivity of tests (experimental design), usability (context/utility), and the definition of refer-



Participants at the ISCA workshop in Bochum

Summer 2003 elsnet The off-workshop event led us to Zeche Zollverein, a UNESCO world heritage site located in Essen. This coal-mine closed in 1986 after a long history of mining. After the impressive guided tour, we were offered excellent food at the local Casino Restaurant.

On the third workshop day, the quality of speech technology was elaborated, including evaluations of textto-speech (TTS) and spoken dialogue systems, and an exploration of parallels in the concepts of sound design and usability engineering.

All in all, this first attempt at getting researchers of

various fields together has offered the opportunity to extend one's view on the auditory quality issue, and to develop new ideas for one's own research work. We hope that further workshops on this topic will follow.

FOR INFORMATION

Florian Hammer is a junior researcher at the Telecommunications Research Center Vienna (ftw)

Email: Hammer@ftwat Web: www.ftw.at

ISCA web site: www.isca-speech.org **Virtools:** www.virtools.com

Announcement

LangTech 2003: Europe's Language Technology and Industry event 24-25 November 2003 Méridien Montparnasse Hotel, Paris

LangTech 2003 is the second instalment of the European forum for language technologies.

Offering a unique platform for the language and speech technology community, LangTech 2003 will bring developers and entrepreneurs together with integrators, investors, and users.

The three main areas covered at LangTech 2003 are: Speech Technologies and Applications; Semantic Web and Knowledge Management; and Multilingualityrelated Solutions, Technologies, or Services.

Keynotes and presentations will be given by industrial key players to an audience made of representatives from major industries participating in the Human Language Technology (HLT) market and venture capital providers. Panels dealing with industrial needs, market trends, users, and research and development for the future will draw the attention of the wider community to topics and issues of significant interest for the promotion and the growth of the language technologies market.

LangTech 2003 will combine these sessions with an exhibition, where companies will be able to showcase their products and services, meet current and potential clients, and promote their activities. In addition, start-ups and SMEs involved in the language technology sector will have the opportunity to introduce them-

selves and promote and pitch their activities, with the objective of attracting investors and clients If you are interested in exhibiting or making a presentation during these 'elevator pitch sessions', you should contact the organisers (see below).

The first European forum for language technologies, LangTech 2002, took place in Berlin last year (see report in *ELSNews 11.4*). It was attended by some 330 participants from over 30 countries and across five continents. The programme featured presentations given by representatives of over 70 companies, from 20 nations, and also included keynotes from major industry players (Bill Dolan, head of the Natural Language Processing unit at Microsoft Corporation, and Wofgang Wahlster from the German Research Centre, DFKI). The exhibition at LangTech 2002 attracted 20 European companies, and 23 SMEs gave presentations during the dedicated sessions.

On-line registration forms for exhibitors and attendees at LangTech 2003 can be found on the web site.

FOR INFORMATION

Email: langtech2003@elda.fr

Web: www.lang-tech.org



SIGDial

Combining dialogue system development with information extraction techniques

Arne Jönsson, NLPLAB, IDA, Linköping University

Dialogue systems are normally developed to access structured data, often stored in databases. However, most information available in electronic for mats is not found in databases; the vast majority comes as text, making up huge sets of unstructured information in natural language. Information extraction techniques that extract relevant information from textual documents can be used to compile such information into a database. The challenge is to combine these areas of language technology research and develop dialogue systems that can access information from unstructured text documents.

In principle this sounds like a straightforward endeavour but, in practice, it involves a number of research issues such as the type of information to extract, handling information gaps, and inference both inside the dialogue and in the interpretation of source documents. We must, for instance, construct a shared domain ontology that captures different conceptualisations of the domain, the one present in the information source and the one users have. Automatic information extraction also means that we must rely solely on the information in the text document, which is often incomplete or expressed in terms that make extraction difficult.

At the Natural Language Processing Laboratory at Linköping University, we have addressed this combined research issue of utilising information extraction techniques to automatically create structured databases from unstructured documents to be accessed by dialogue systems. Our first such system, BirdQuest, was developed based on a bird encyclopædia from which information was extracted and transformed to a relational database. The interaction component was developed from a framework for dialogue systems development [1].

The BirdQuest ontology was developed from two different types of empirical material: the bird encyclopædia and a question corpus collected on a web site for a Nature programme on Swedish television. From the encyclopædia a conceptualisation underlying the structure and presentation of information to be used by the information extraction was constructed. The result was a system-oriented domain ontology representing experts' view of the domain. The question corpus yielded a user-oriented conceptualisation of the domain, thus providing a non-expert view of the domain useful for the interaction component. These two



conceptualisations were then merged to form a shared domain ontology

for all components of the system. Taking the system-oriented ontology as a starting point, new categories found in the question corpus were added. Allowing multiple inheritance, new links between existing categories and new categories were added.

We evaluated BirdQuest in a small public study with 27 users, having no knowledge of dialogue systems or interest in birds, and without any specific instructions [2]. BirdQuest correctly interpreted 48% of the users' utterances. Many of the utterances, 23%, were out of coverage, such as "How do you kill crows?", unfortunately not unusual when the public is invited to try out dialogue systems. BirdQuest also often failed because of the ontology being incomplete. Many concepts are hard to capture. One notable problem is colour. A bird is not described as having one colour. Instead each body part has its own colour and many descriptions are provided in terms of other birds

To sum up, for certain types of information it is possible to utilise information extraction techniques to create databases to be accessed by dialogue systems. When developing such systems we are faced with new challenges such as capturing a shared view of the information and handling information sources written for the purpose of being read and understood by humans and not computers.

FOR INFORMATION

Arne Jönsson is associate professor in the NLPLAB at Linköping University

Email: amejo@ida.liu.se

Web: www.ida.liu.se/~arnjo

References [1] nlpfarm.sourceforge.net

[2] Annika Flycht-Eriksson and Arne Jönsson, Some empirical findings on dialogue management and domain ontologies in dialogue systems – Implications from an evaluation of BirdQuest, 4th Annual SIGdial Workshop on Discourse and Dialogue, Sapporo, 2003.





ß

CL2003: the International Conference on Corpus Linguistics

Andrew Roberts, University of Leeds

After the initial success of the first Corpus Linguistics conference in 2001 in honour of Geoffrey Leech's 65th birthday, Tony McEnery et al. decided to make it a more permanent fixture in the linguistics calendar, and stage the conference every two years. Hence, it was time for the sequel to commence, and, as a result, hundreds of corpus linguists were seen flocking to Lancaster. The peaceful university campus, located out of town, with its countryside atmosphere, was a lovely setting for the event. Luckily, the weather also blessed us with fine sunshine throughout – much to the relief of the many international participants who were expecting the infamous British climate (i.e., wet and cold!).

Around this time two years ago, Geoffrey Sampson wrote an article in *ELSNews* regarding the future role of ICAME (in particular, their conferences). He remarked that ICAME's focus on English research, coupled with its restrictions on conference sizes to ensure a *friendly* atmosphere, means that not only does it miss out on the ever increasing (and exciting) shift towards non-English language research, but also the next generation of researchers, with energy and fiesh ideas, but who are not

yet established enough to join the ICAME clique. He went on to comment about the success of CL2001 that had taken place shortly before, and predicted success for its future. ICAME's weaknesses are CL's strengths, which is why Sampson's prediction was correct.

The conference kicked off with a day of workshops covering development of learner corpora and multilingual corpora, corpus-based approaches to figurative language, and shallow processing of large corpora (SProLaC). The following four days saw papers presented in three parallel sessions. With approximately 95 papers and 30 posters on offer, it would be impractical to go into any real depth about them. Needless to say, all areas of the field were well represented. There were reports of new resources and research that is being developed for European minority languages and South-East Asian languages. The usual suspects, such as tagging, parsing disambiguation, grammars, and information extraction were well covered. Unsurprisingly, there was also a strong focus on corpus development, annotation, and tools. Even if translation studies, exploiting corpora, and semantics are added to the list, it is still



Corpus Linguistics researchers from Leeds University, alongside Geoffrey Leech (far right), original raison d'être for the Corpus Linguistics conference. And rew Roberts is on the far left.

not exhaustive – which simply illustrates the breadth of the conference. Of course, the fact that each piece of research presented was, by definition, linked to corpora, means that there was a common thread throughout, and so, as varied as the topics were, they never felt disjointed.

Invited speakers were Michael Hoey, Nancy Ide, Susan Hunston, Geoffrey Sampson, and Nicoletta Calzolari. All are well known within the field and offered fascinating and well received talks. Probably the most infamous presentation of the conference was by Tony McEnery, covering his studies of 'the f-word' within the BNC. It was even rumoured that Eric Atwell, a speaker in a parallel session, was recommending to his audience to go and hear Tony's instead as it was more interesting!

The hospitality was excellent throughout the conference, especially the catering, which was of a particularly high standard. Naturally, coffee breaks and meal times were the main social occasions. None more so than on the third day, when we were all whisked a way by coach to the magnificent Ashton Memorial within the beautiful Williamson Park, which sits high above the main town centre. From here, people could enjoy the lovely scenery, most notably the mountains of the Lake District National Park. The memorial building itself wasn't particularly large, and with so many people in it, personal space was quickly becoming a luxury. If there was anybody with whom you weren't acquainted, you were by the end of the night, which was a good thing in my opinion!

The conference was very well organised, and was probably as close to optimal as you could get. Whilst there were approximately 200 participants (representing 30 countries), the atmosphere was still very friendly and informal. Also, three does appear to be the magic number in terms of the number of parallel sessions. At times, choosing one of three talks could be a difficult

contd from p.11

the assistance of Paola Baroni and Monica Monachini.

One of the objectives of this Workshop is to launch the ICCWLRE (International Co-ordination Committee for Written Language Resources and Evaluation).

The Workshop, originally planned as an ACL 2003 Workshop, will be held in Paris on 28th and 29th August 2003.

Further information about both the ENABLER Network and the ENABLER/ELSNET Workshop can be found at the website.

decision, but at least it gives a degree of flexibility. Any more and I think that people may begin getting frustrated at missing too many talks, especially when there is more than one presentation of interest at a given time – as is often the case at some of the larger conferences. Therefore, congratulations should go to Tony McEnery, Dawn Archer, Paul Rayson, Andrew Wilson, plus the many other local staff who helped to ensure an enjoyable and interesting conference. We look forward to CL2005!

FOR INFORMATION

Andy Roberts is a research student at the University of Leeds

Email: andyr@comp.leeds.ac.uk

Web: www.comp.leeds.ac.uk/andyr

Proceedings of this and previous CL conferences and the SProLaC workshop are available as UCREL technical papers as follows:

2003: Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery (eds.). Proceedings of the Corpus Linguistics 2003 conference. UCREL technical paper number 16. UCREL, Lancaster University

2001: Paul Rayson, Andrew Wilson, Ton y McEner y, Andrew Hardie and Shereen Khoja (eds). Proceedings of the Corpus Linguistics 2001 conference. UCREL technical paper number 13. UCREL, Lancaster University

SProLaC: Kiril Simov and Petya Osenova (eds) (2003). Proceedings of the The Workshop on Shallow Processing of Large Corpora (SProLaC 2003) held in conjunction with the Corpus Linguistics 2003 conference. UCREL technical paper number 17. UCREL, Lancaster University.

FOR INFORMATION

ENABLER

Web: www.enabler-network.org

Wor kshop

Email:

Nicoletta Calzolari: glottolo@ilc.cnr.it Alessandro Lenci: alessandro.lenci@ilc.cnr.it Steven Krauwer: steven.krauwer@elsnet.org Paola Baroni: eagles@ilc.cnr.it Monica Monachini: monica.monachini@ilc.cnr.it Summer 2003



New members

More new ELSNET members!

Centre for Language Techology Macquarie University Sydney, Australia (*Robert Dale*)

Yes, you are reading *ELSNews*, and yes, a research lab that's about as far away from Europe as you can get has joined ELSNET. This is not as odd as it sounds: many of the staff in our lab have connections with Europe, but, more importantly, Australia is just about to embark on the creation of a number of research networks similar in spirit to ELSNET, and we aim to build bridges to ELSNET and other international networks.

The Centre for Language Technology has been around under its current name since 2001, but has a history going back to 1994 when we were a Microsoft-funded research lab. Our research is built around a number of key projects that cover a wide range of issues in speech and language processing, including:



The CIT building

XtraMind Technologies, GmbH, Saarbrücken, Germany (*Klaus Netter*)

XtraMind Technologies GmbH is a provider of intelligent software solutions for the optimisation of electronic customer communication in business. Based on advanced methods of Artificial Intelligence and Language Technology, the solutions by XtraMind support and automate individual customer dialogue over communication channels such as e-mail, digital fax, and the web.

XtraMind's product portfolio comprises standard software solutions, such as *XM-MailMinder*, a leading solution for professional E-Mail Response Management covering the full life-cycle of electronic communication in service centres and enterprises. This application

- PENG, a controlled language and associated tools
- AnswerFinder, an answer-extraction system
- Coral, a natural language generation system for route descriptions
- KES, a system that integrates text categorisation, information extraction, and text summarisation
- FON, a system for handling meeting-room speech

You can find out more about these projects at our web site. The CLT collaborates closely with CSIRO, the Australian Government research organisation, and is involved in the Capital Markets Co-operative Research Centre, which aims to apply language technologies in the financial domain.

We're always happy to receive visitors, so if you're thinking of spending your sabbatical in Sydney, let us know!



Sydney Harbour Bridge

FOR INFORMATION

Director of the Centre and ELSNET contact: Professor Robert Dale Division of ICS Sydney NSW 2109 Australia Email: Robert.Dale@mq.edu.au Tel: +61 413 383 248 Web: www.clt.mq.edu.au

is based on a technology platform *XM-MindSet*, a suite of intelligent software components for the multilingual analysis and processing of natural language content.

XM-MindSet components are also integral parts of individual advanced solutions developed by XtraMind and its partners.

XtraMind was founded in 2000 as a spin-off of the German Research Center for Artificial Intelligence (DFKI), which for the first time in history became a shareholder of a spin-off



Dr Klaus Netter



company. Located in the DFKI-building on the campus of the University of the Saarland, XtraMind now has more than 50 employees and has established itself in the market with reference customers such as (among others) 1&1 Internet, Blaupunkt, Bosch Communication Center, Deutsche Bahn, GMX, ProSieben.Sat.1, Quelle, and ratiopharm. At the level of sales and services XtraMind is cooperating with partners such as Materna, Siemens Fujit su Corporation, Cambridge Technology Partner, twenty4help, and caatoosee.

FOR INFORMATION

ELSNET contact: Dr. Klaus Netter

XtraMind Technologies GmbH Stuhlsatzenhausweg 3 D-66123 Saarbrücken Ger many

Tel: +49 681 302 5100 **Fax:** +49 681 302-5109 **Web:** www.xtramind.com **Email:** netter@xtramind.de

ScanSoft, Aachen, Germany (*Rainer* Siemund)

ScanSoft is a publidy traded company (Nasdaq: SSFT) with offices all over the world, that develops imaging, speech, and language solutions. Its imaging solutions include document automation solutions such as OCR (optical character recognition), eForm design, and personal document management applications. Its speech and language technology solutions include speech technology for document creation (Dictation), technologies that enable the voice-control of computer systems (ASR) for telephony, PC/multimedia, and for embedded systems, and technology that allows one to add humansounding synthesised voice to software applications and embedded hardware systems (TTS).

ScanSoft products include imaging solutions such as *Omnipage*, *Omniform*, and *Paperport*, and speech technologies such as the *RealSpeak* class of TTS systems, the *SpeechPearl* class of ASR systems, and the document creation product *Dragon NaturallySpeak ing*.

ScanSoft employs almost 500 persons worldwide.

ScanSoft Aachen's research focusses on improving ASR beyond the current state-of-the art and increasing the naturalness, especially for the teleph-



XtraMind employees enjoying a casual chat on the stairs



ony server-based and the automotive markets, and more generally for incorporation in embedded systems.

Scansoft, and the companies it has recently acquired, have been involved in a wide range of national and European research and collaboration projects, including SpeechDat, SpeechDat-Car, SpeechDat-E, Interface, ISLE, and others. It is currently a participant of the EU-funded projects SPEECON, Orientel, and NICE.

FOR INFORMATION

ELSNET contact: Dr Rainer Siemund

Kackertstrasse 10 52072 Aachen Germany

Tel.: + 49 241 8871 0 **Fax:** + 49 241 8871 140 **Web:** www.scansoft.com **Email:** rainersiemund@scansoft.com Summer 2003



Opinion

Will Chomskyan linguistics be googled out of existence?

Annie Zaenen, Xerox PARC

While in the eighties linguists and computational linguists collaborated on the elaboration of parsers and other tools for natural language analysis, the nineties saw a disengagement on both sides (or, maybe I should say, a disengagement on the computer scientists' side as on the linguistic side the interest had always been that of a minority deemed worthy of the attention of mainstream linguists only to be condemned). Computational methods of natural language analysis moved from the symbolic approaches shared with linguistics to mainly statistical ones, a change reflected in that of the preferred name of the field from Computational Linguistics to Natural or Human Language Processing. Currently courses in natural language processing are mainly taken by students in computer science and in statistics and have by and large moved out of linguistics departments. Some departments offer language engineering courses which in the nineties seemed to be a good way to ensure some more job opportunities for students but which have little theoretical impact.

In the last couple of years, however, odd happenings can be observed: linguistics conferences do have special sessions where statistical methods are discussed, not just in the context of sociolinguistics where they have always been in fashion but also in connection with theoretical syntax and, even more, phonology. Within phonology the impulse came mostly from phonetics, within syntax it comes from the revival of corpus linguistics. This revival is technologically driven: it is the availability of linguistic data in electronic form, and especially that of parsed and annotated corpora, that is pushing towards a revolution in syntactic methodology and in what is considered to be the scope of the field.

The easy availability of corpus data allows linguists to better study phenomena that have always been recognised as difficult to investigate via introspection, mainly those that are not a question of grammaticality or not but that appeal to more subtle distinctions in acceptability. (In that respect corpus linguistics is of course also a great boon to foreign language students.) The effects, however, go further. Together with the vogue of Optimality Theory, corpus linguistics has lead to a renewed interest in syntactic variation within and across languages. This in turn leads syntacticians to become interested in statistical methods.

Of course statistics have played a role in linguistics before, enough to be ridiculed by Chomsky in the fifties and the sixties. And when computers first became available for such tasks in the sixties, some linguists started to use corpora to count. But there is a difference. This time the statistics are about the objects of syntactic investigation: abstract patterns, not surface strings or lexical items. And the statistical methods used are not open to the facile attacks that Chomsky launch ed earlier. Even when he wrote those, there were already more sophisticated statistical models than the ones he talked about, but the field has made great progress since (see Pereira, 2000, for an overview[1]).

For better or worse, linguists have always had a habit of exploiting new techniques developed in other disciplines to find new insights in language. This is happening again. There are now courses in theoretical syntax, both in linguistics departments and in computer science departments, where statistical methods are used to investigate the distribution of syntactic constructions as well as lexical variation. As the data for these studies is only manageable in electronic form, the courses also include an introduction to basic computational tools.

The impact of corpus linguistics, however, is more insidious than just helping linguists look at the distribution of syntactic structures. Even a cursory use of the Internet for such purposes leads to worrying questions about grammaticality judgements. Not because we find examples that do not confirm those found in textbooks: this is to be expected by those who believe in a distinction between competence and performance. But because once one is confronted with these examples, even the most stringent grammarian cannot find anything wrong with them - they cannot be written off as 'performance errors', or a difference between idiolects. A case in point is discussed in a talk by Chris Manning at the 2001 AAAS Symposium on Mathematical Statistics in Natural Language Analysis[2]. He gives the following examples from the New York Times of the use of "consider as":

The boys consider her as family and she participates in every - thing we do

Green.span said, "I don't consider it as something that gives me great wnem".

"We consider that as part of the job," Keep said.

Although the Raiders missed the playoffs for the second time in the past three seasons, he said he considers them as having champions hip potential.

Culturally, the Croats consider themselves as belonging to the "civilized" West, ...



For those who don't remember. "consider as" is consid-

ered to be ungrammatical. Similar examples can be multiplied for any subcategorization constraint proposed. These results lead to a relativisation of the notion of grammaticality judgement itself. What if it is just a form of linguistic performance among others? And moreover one that is often influenced by what is common within the current world rather than what is linguistically possible?

We seem not far away from a return to more empirically based theorising, something that everybody who has witnessed the Chomsky-driven excesses of idle speculation that have fuelled much of theoretical syntax over the last thirty years has to applaud. In fact we might be witnessing a paradigm shift towards syntactic models that are inherently probabilistic. Of course, this will not happen overnight and not without one of the Picrocholine wars[3] that linguists relish so much. A shot was already fired at the last LSA where the president made an uncharacteristic use of his privilege to address the full audience by chiding part of it for, in his opinion, confusing use and grammar[4].

However these debates may turn out and although the aims of theoretical and computational linguistics will remain distinct, as they have always been, the methodological chasm between the two is in the process of being narrowed. If you haven't set foot in your local linguistics department for the last ten years, drop by one of these days, you might run into a couple of theoretical syntacticians that at the very least know what a log-linear model is. It is not likely, though, that they will be very interested in Information Retrieval.

FOR INFORMATION

Annie Zaenen is Principal Research Scientist at Xerox PARC, USA

Email: zaenen@parc.com Web: www2.parc.com/istl/members/zaenen

References

[1] Fernando Pereira. Formal grammar and information theory: Together again? Philosophical Transactions of the Royal Society, 358(1769):1239-1253, April 2000. [2] Chris Manning. Frequencies and Probabilities within the Grammars of Natural Languages, San Francisco, 2001 [3] Rabelais. Gargantua.

[4]Frederick J. Newmeyer. Grammar is Grammar and Usage is Usage, presidential address, LSA. Atlanta, January 2003

The ENABLER Thematic Network

The ENABLER Thematic Network, launched in the framework of the EU Project "ENABLER -European Network Activities for Basic LanguagE Resources" (Contract Number: IST-2000-31069), aims at improving cooperation among national activities established by national authorities for providing Language Resources (LRs) for their languages.

The action aims at:

- establishing a regular exchange of information;
- identifying and fostering possible synergies and cooperation;
- promoting the compatibility and interoperability of their results, thus facilitating the successful transfer of technologies and tools among languages and the construction of multilingual LRs;
- increasing the visibility and the strategic impact of those national activities in the field of HLT;

• contributing to the creation of an overall framework in which the public and private sectors, national efforts, and international coordination could cooperate in order to answer the IST need for LRs.

The ENABLER Network members are: Università degli Studi di Pisa (I) Institute for Language and Speech Processing (EL) European Language Resources-Distribution Agency (F) Center for Sprogteknologi (DK)

Nederlandse Taalunie (NL)

Institut d'Estudis Catalans (E)

Ludwig Maximilians Universität München (D)

Fundação da Universidade de Lisboa - Centro de

Linguística da Universidade de Lisboa (P) Faculty of Mathematics and Physics - Charles University in Prague (CZ)

Beltext – Université de Liège (B)

Department of Speech, Music and Hearing (CH)

Real Academia Española (E) Instituut voor Nederlandse Lexicologie (NL)

The first named member is the ENABLER Project Coordinator; the second, third and fourth are the ENABLER Project Principal Contractors; the remaining named institutions are the ENABLER Project Members.

A Workshop entitled "International Roadmap for Language Resources" is being organised under the auspices of ENABLER and ELSNET by Nicoletta Calzolari, Alessandro Lenci, and Steven Krauwer with

contd on p.7

Summer 2003

Announcement



ParGram: Developing Parallel Grammars

Helge Dyvik, University of Bergen

Helge Dyvik describes the ParGram project – a multisite project that aims to de velop parallel large-scale grammars of several dif-ferent languages.

The Point of Large-Scale Grammars

Within Natural Language Processing, rule-based methods of 'deep' grammatical and semantic analysis on the one hand, and statistics-based machine learning methods on the other, are traditional competitors, but in recent years they are increasingly seen as complementary approaches that can be combined in innovative ways. This makes broad-coverage grammars which perform deep syntactic and semantic analysis and can be efficiently processed by parsers and generators a desirable part of the basic resources a language needs for its language technology. The general development of hardware capacity, together with new insights in the parsing of complex grammars, have also infused large-scale grammar development with renewed interest.

One of the advantages of large-scale grammar development is that it enforces consistency across the descriptions of different grammatical phenomena in a language. While isolated grammar fragments dealing with limited subsets of the language often turn out to be incompatible if combined, a formalised large-scale grammar will ultimately have to deal with all the interactions among the various constructions and lexical entries of the language. Hence, as a resource grammar it will support the mutual compatibility of possible smaller, application-specific grammars derived from it.

An even more demanding ambition is consistency across languages. Such consistency will probably facilitate machine translation and other kinds of multilingual language processing. However, even agreeing on a common general framework for grammatical description is not enough to ensure a desirable degree of cross-linguistic consistency. Differences between the grammars developed for different languages will just as often reflect a typology of linguists and their preferences as one of languages, even when a common framework is applied. This makes close co-operation during the grammar development process desirable.





The Grammars and the Platform

ParGram - the Parallel Grammar Project - is a longterm project aimed at the development of large-scale computational grammars for several languages, using the same evolving grammar engineering platform and based on the same principles of grammatical description. The central site for ParGram is the Palo Alto Research Center (PARC, formerly Xerox PARC) in California, while other collaborating groups are located in academic and corporate research institutions in a number of countries ParGram started in 1993 with the three languages English, French, and German, and has later been joined by groups working on Norwegian, Japanese, Urdu, and Korean, with further languages appearing on the horizon as possible future members. The largest ParGram grammars are the grammars of English and German, which have been developed and tested on the basis of extensive text corpora. The English grammar provides full syntactic analyses for 74.7% of the sentences in the one-million-word UPenn Wall Street Journal corpus, and partial analyses for the remaining 25.3% [1]. Covering corpora like this necessitates the development of linguistic analyses of many constructions that are not normally discussed in the linguistics literature.

The development platform of ParGram is the Xerox Linguistic Environment (XLE), developed at PARC. XLE is a tool for practical grammar engineering which incorporates efficient parsing and generation algorithms



and allows the same grammars to be used for parsing and generation. The platform implements the syntactic theory Lexical Functional Grammar (LFG), one of the eadiest unification-based syntactic frameworks, dating from the late seventies, whose formalism and basic architecture has remained comparatively stable through the years. The interface to XLE is closely modelled on the older Interlisp-based Grammar Writer's Workbench of Xerox PARC, but the system itself, as a platform for grammar engineering, is designed to handle industrialsize grammars and lexicons in flexible and efficient ways. An important purpose of ParGram is to let the continuous development of the XLE platform be informed by the needs of the grammar writers and the peculiarities of the languages under description.

Achieving Parallellism

Another important purpose of the project is the achievement of parallel grammars. This is primarily taken to mean parallellism on the level of 'f-structure', one of the two basic levels of syntactic representation in LFG. The f-structure of a sentence, a re-entrant attribute-value matrix, represents its basic predicate-argument structure linked to syntactic functions like subject and object, and furthermore its constellation of grammatical features expressing such categories as number, case, tense etc., while abstracting away from word order, among other things The assumption of LFG is that there is much less cross-linguistic variation on the f-structure level than on the level of c-structure, the latter representing the organisation of the sentence in phrases and subphrases in the form of a phrase-structure tree. However, the aim of ParGram is not to achieve near-identical f-structures for translationally corresponding sentences, glossing over cross-linguistic variation. Rather, it is to describe the same grammatical phenomena across languages in the same way on the fstructure level, using the same inventory of features and values, while at the same time respecting the typological differences among the languages.

The procedure to achieve this involves meetings twice a year, parts of which are dedicated to detailed discussions of selected grammatical phenomena and their description in the various grammars. The aim is always to reach consensus on descriptive principles and the inventory of relevant grammatical features with their possible values. Differences between the grammars are acceptable if and only if it can be argued that they reflect differences between the languages and not just different preferences among the grammar writers. This is a highly demanding and extremely interesting process, where broad and detailed linguistic insights in form and meaning must be coupled with an understanding of the mathematical underpinnings of grammatical formalisation and considerations of efficient processing.

Extending the Platform

Another topic at the meetings concerns desirable further developments of the XLE platform, relating to such properties as e.g., the handling of special linguistic phenomena, robust and efficient parsing, and general userfriendliness of the interface. An example of the first of these is the analysis of so-called complex predicates in Urdu. The analysis of complex predicates involves describing in the syntax certain semantic phenomena which European languages mostly seem happy to treat in the lexicon, a fact which was originally reflected in the architectures of LFG and XLE. The Urdu data spurred further development of the platform to enable it to treat the phenomenon in a motivated way, which in turn led to a reconsideration of certain analyses in some of the other grammars as well. Since XLE supports the full syntactic functionality of the LFG theory, developments like this may at the same time contribute to the discourse within the theoretical linguistic community.

Robustness and efficiency are important considerations in the development of XLE. One implemented device is Optimality Marking, which gives a handle on the treatment of ambiguity by allowing a ranking of competing analyses for ambiguous expressions The device is also used in the service of robustness to allow 'dunk' or 'fragment' parsing, i.e., finding the analyses of the maximal analysable chunks of sentences that are not completely covered by the grammar. By ranking a rule set for fragment parsing below the regular rules, fragment parsing will always be tried as a second option when full parsing fails. With regard to efficiency, XLE incorporates various devices for limiting the amount of work done by the parser, allowing performance to approach linear or even constant time at controlled costs. Incorporation of statistical approaches to disambiguation is also possible, as well as combinations with various 'shallow' methods, such as parsing text which has been marked up with part-of-speech and named entity tagging This creates a laboratory for exploring the relative merits of deep and shallow methods. Experience from handling the large ParGram grammars is essential in giving direction to these extensions of the platform.

Why Theory is Practical

The advantages of basing grammar development on well-reasoned linguistic principles and models become abundantly dear as the grammars grow. In general, the addition of new construction types to a grammar takes less and less effort, since their proper interactions with existing constructions follow more or less automatically as long as the latter are described in a principled and perspicuous way.

A related consideration concerns the addition of new languages to the project. In the first place, any new

Summer 2003 *elsnet* grammar-writing group will have a body of tested and agreed-on features and principles to start from. But in the second place attempts are also made to mould a grammar for a new language from an existing grammar for a (typologically) closely-related language, thereby minimising the development effort. Such attempts are being made for Korean based on the Japanese grammar, and for Danish based on the Norwegian grammar Attempts like these clearly presuppose that the existing grammars are linguistically perspicuous

Applications of the Resource Grammars

A resource grammar should ideally be useful in a variety of different applications, and maximally independent of specific theoretical choices made for those applications. This might appear to be a drawback for a grammar based on a specific linguistic theory, like the ParGram grammars. Still, a 'theory-neutral grammar' is not a coherent concept, and hence no alternative. Rather, the ambition must be to use a theoretical framework which allows communication with other frameworks as flexibly as possible.

The so-called 'projection architecture' of LFG is interesting in this connection. The representations assigned to expressions by an LFG grammar are derived by codescription. This means that one representation is not derived by processing another, rather, the representations are brought into systematic correspondence with each other by associating partial descriptions of one with the rules deriving the other. In this way any number of new representations for sentences can be added to the grammar by 'projecting them off' existing representations

This facility is being explored in the Norwegian grammar, which is now a resource for a Norwegian project called LOGON, concerned with transfer-based translation from Norwegian into English. The English target grammar is written in the HPSG framework and uses representations based on Minimal Recursion Semantics (MRS) as its semantic interface. MRS representations bear no direct relation to the representations of Lexical-Functional Grammar, but the LFG architecture still makes it comparatively easy to project MRS representations off the existing structures, exploiting the information already there. Thus, this possibility facilitates the integration of the resource grammar into applications using different formalisms.



and their condensations.

Conclusion

The ParGram project is an ambitious attempt to combine broad-coverage, deep grammatical analysis with multilingual parallellism, efficient processing, and other language processing methods. In addition to providing useful resource grammars for a number of languages,

project the also contributes new insights into natural language parsing and generation, at the same time documenting the viability of large-scale grammar development as part of the repertoire of approaches Natural to Language Processing.



Helge Dyvik

FOR INFORMATION

Helge Dyvik is Professor of General Linguistics at the University of Bergen, Norway

Email: helge.dyvik@lili.uib.no **Web:** www.hf.uib.no/LiLi/SLF/seksjonsleder

ParGram project web sites: Main page: www.parc.xerox.com/istl/groups/nltt/pargram Stuttgart: www.ims.uni-stuttgart.de/projekte/pargram/ Fuji Xerox: www.fujixerox.cojp/crc/rf-kndg/linguistic/ Bergen: www.ling.uib.no/~victoria/NorGram/ UMIST: www.ccl.umist.ac.uk/staf f/mutt/pargram/

Reference

[1] Riezler, Stefan, Tracy H. King, Ronald M. Kaplan, Richard Crouch, John T. Maxwell, III, and Mark Johnson. 2002. Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques In *Praceedings of the 40th Annual Meeting of the ACL*, Philadelphia.

P

Calendar

Future Events

| Aug 10 | <i>Third IJCAI workshop on Knowledge and Reasoning in Practical Dialogue Systems</i> : Acapulco, Mexico Email: ijcaiws@ida.liu.se URL: www.ida.liu.se/~nlplab/ijcai-ws-03 | | | | | | |
|------------|---|--|--|--|--|--|--|
| Aug 18-29 | 15th European Summer School in Logic, Language and Information: Vienna, Austria Email: esslli03@logic.at URL: www.logic.at/esslli03 | | | | | | |
| Aug 27-29 | ISCA Workshop on Voice Quality: Functions, Analysis and Synthesis: Geneva, Switzerland Email: Christophe.D'Alessandro@limsi.fr URL: www.limsi.fr/VOQUAL | | | | | | |
| Aug 28-31 | ISCA Workshop on Error Handling in Spoken Dialogue Systems: Chateau-d'Oex-Vaud, Switzerland Email: errorworkshop@speech.kth.se URL: www.speech.kth.se/error | | | | | | |
| Sept 1-4 | Eurospeech 2003/Interspeech 2003: Geneva, Switzerland Email: organisers@eurospeech2003.org URL: www.eurospeech2003.org | | | | | | |
| Sept 4-6 | DiaBruck 2003 (Seventh Work shop on the Semantics and Progratics of Dialogue): Wallerfangen, GermanyEmail: diabruck@coli.uni-sb.deURL: www.coli.uni-sb.de/diabruck | | | | | | |
| Sept 8-9 | Speech Processing Workshop (in connection with DAGM03): Magdeburg, Germany Email: wendemu@ipe.et.uni-magdeburg.de URL: speech-dagm03.uni-magdeburg.de | | | | | | |
| Sept 10-12 | Recent Advances in Natural Language Processing (RANLP): Borovets, BulgariaEmail: ranlp03@lml.bas.bgURL: www.lml.bas.bg/ranlp03 | | | | | | |
| Sept 23-28 | Machine Translation Summit IX: New Orleans, USA Email: focalpoint@amtaweb.org URL: www.mt-summit.org | | | | | | |
| Sept 25-26 | First Nordic Symposium on Multimodal Communication: Copenhagen, DenmarkEmail: cst@cst.dkURL: www.cst.dk/mumin | | | | | | |
| Oct 22-25 | Fifth International Workshop on Multidisciplinary approaches to Discourse: Driebergen, NetherlandsEmail: l.lagerwerf@scwvu.nlURL: home.scw.vu.nl/~lagerwerf/Mad03Web | | | | | | |

Submission deadlines

| Aug 1 | TLT2003: Växjö, Sweden, Nov 14-15, URL: www.msi.vxu.se/~rics/TLT2003 | | | |
|----------------|---|--|--|--|
| Sept 1 | CoLogNET-ELSNET Symposium: Amsterdam, Netherlands, Dec 18, URL: www-uilots.let.uu.nl/~ctl/workshops/CES03 | | | |
| Sept 30 | CULT-BCN 2004: Barcelona, Spain, Jan 23-25, URL:www.fti.uab.es/cg.cult.bcn | | | |
| Oct 1 | Machine Learning Journal: Special Issue on Learning in Speech and Language Techologies, URL: www.ee.ust.hk/~pascale/MLJspecial.html | | | |
| Oct 1 | Journal of Computer Speech and Language: Special Issue on Word Sense Disambiguation, Email:Judita.Preiss@cl.cab.ac.uk | | | |
| This is only a | selection – see www.elsnet.org/cgi-bin/elsnet/events.pl for details of more events and | | | |

deadlines.pl for more deadlines.

If you would like to write a review of any of these (or other language/speech related events you attend), please contact the *ELSNews* editor.

Summer 2003



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

Task Groups

16

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

Linguistic & Speech Resour of Antonio Zampolli, Istituto di Linguistica Computazionale (I) and Ulrich Heid, Stuttg art University (D)

Research

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

Executive Board

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

| | The ELSNET Participants: | | I | IRST | D | Langenscheidt KG |
|--------|--------------------------|--|-----------------|--|----------|--|
| | | | Ι | Consorzio Pisa Ricerche | D | Verlag Moritz Diesterw eg GmbH |
| | Acauci | inc sites | Ι | Fondazione Ugo Bordoni | D | aspect Gesellschaft für Mensch-Maschine |
| | | | Ι | Consiglio Nazion ale delle Ricerche | | KommunikationmbH |
| | A | University of Vienna | I | Università degli Studi di Pisa | D | Philips Research Laboratories |
| | A | A ustrian Research Institute for Artificial | IRL | Trinity College University of Dublin | D | Grundig Professional Electronics Gmb H |
| | | Intelligence (OFAI) | IRL | University College Dublin | D | Acolada Gmbh |
| | А | Vienna University of Technology | IT | Inst of Mathematics & Informatics | D | IBM D cutschland |
| | AUS | Macqua rie University | NI | Leiden University | D | Varetis Communications |
| | В | Katholieke Universite it Leuven | NL | Foundation for Speech Technology | D | XtraMind Technologies Gmb H |
| | В | University of Antwerp - UIA | NL | University of Twente | D | Scansoft Aachen G mbH |
| | BG | Bulg Acad. Sci Institute of Mathematics | NI | Tilburg University | D | Heartsome Europe GmbH |
| 、 、 | | and Informatics | NI | Technische Universiteit Findhoven | D | Sympolog Voice Solutions GmbH |
|) | BY | Belor ussian Academy of Sciences | NI | University of Amsterdam (UvA) | DK | Tele Dan mark |
| | CH | SUPSI University of Applied Sciences | NI | Netherlands Organization for Applied | DK | Zaco A/S |
| | CH | University of Geneva | INL | Scientific Research TNO | E | SchlumbererSema sæ |
| | CZ | Charles University | хл | Scentic Research TWO | E | Telefonica L& D |
| echt | D | In stitut für Angewandte | NL | Utrecht University | E | KNOWI ECDE SA |
| | | Informations for schung | NL | University of Nijmegen (KUN) | EL | LINGA sort |
| | D | Universität Erlangen-Nürnberg - FORWISS | NL | University of Groningen | F | LINGA S.d.I.I. |
|) | D | Universitaet des Saarlandes | NO | Norwegian University of Science and | F | Systian SA New Present Centre Frances |
| | D | German Research Center for Artificial | | Technology | Г | Mamadata |
| r ces | | Intelligence (DFKI) | NO | University of Beigen | Г | America |
| | D | Christian-Albrechts University, Kiel | P | University of Lisbon | Г | Aelospanale |
| | D | Universität Hamburg | P | IN ESC ID Lisboa | Г | VECSIS |
| l | D | Univer sität Stuttgart-IMS | P | New University of Lisbon | Г | SCIPER |
| | D | Ruhr-Universität Bochum | PL | Polish Academ y of Sciences | F EIN | IGID Kielikere Ori |
| | D | Universität de s Saarlan des CS-AI | RO | Roma nian Academy | FIN | Nelia Pesserch Center |
| | DK | Center for Sprogteknolo gi | RU | Russian Academy of Sciences, Moscow | LU1 | Normhal a gial tal |
| 5 | DK | University of Southern Den mark | S | KTH (Royall institute of Technology) | HU | OL WETTE DICEDCA SC- A |
|) | DK | Aalborg University | S | Linköping University | 1 | |
| · · | Е | Polytechnic University of Catalonia | IR | Sa band University | 1 | TUDE |
| | Е | Universidad Politécnica de Valencia | UA | IRIC UNESCO/IIP | | Compulsor |
| _ | Е | University of Granada | UK | University of Edinburgh | NI | Knowledge Consents PV |
| d | E | Universidad Politécnica de Madrid | UK | Leeds U niversity | NI | Sorbs on NV |
| | E | Universidad Nacion al de Educación a | UK | University of Sheffield | NI | Tarthand P.V |
|) | | Distancia (UNED) | UK | University of Essex | NI | IP Globe he tNederland RV |
| 5, | Е | Universitat Autono ma de Barcelona | UK | University College London The Overset's Linker site of Delfect | DI | Neurosoft Sp. 7.0.0 |
| .) | EL | In stitute for La nguage & Speech Processin g | UK | The Queen's University of Benast | PU | Rearosoft Sp. 20.0 |
| | | (ILSP) | UK | University of Brighton | RU | AN ALIT Itd |
| | EL | National Centre for Scientific Research | UK | University of TORK | S | Sama Infodata |
| | | (NCSR) 'Demokritos' | UK | United to the second se | S | Teliz Promotor AB |
| | EL | University of Patras | | University of Elleter | UK | Vocalis I td |
| | F | Université Paul Sabatier (Toulo use III) | | University of Combridge | UK | Imagination Technologies n.c. |
| | F | LORIA | | University of Sussay | UK | Hewlett-Packard Laboratories |
| .) | F | IRISA/ENSSAT | | University of Sunderland | UK | Canon Research Centre Europe Ltd |
| | F | Université de Provence | UK | University of Sunderland | UK | A IPNET LIK Limited |
| | F | LIMSI/CNRS | Tra directori e | J C 400 | UK | Reuters I td |
| | F | Inst. Nation al Pol ytechnique de Grenoble | maustra | 1 5 1105 | UK | SRI International |
| | GE | Tbilis i State University, Centre on Language, | В | DHaxle y Translations | UK | Sham Laboratories of Furope Ltd |
| | | Logic and Speech | D | No votech GmbH | UK | BT Adactral Park |
| | HU | Lóránd Eötvös University | D | Sympalog Speech Technologies AG | UK | Losica Cambridge Ltd. |
| | HU | Technical University of Budapest | D | DamlerChrysler AG | UK | 20/20 Speech LTD |
| of | | | | | | |

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 academia 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

P