elsnews 9.1

The Newsletter of the European Network in Human Language Technologies Spring 2000

What will you do for ELSNET in FP5?

Steven Krauwer, *ELSNET Coordinator, Utrecht University*

Current members of the ELSNET Executive Board, together with representatives from VDI/VDE and ARAX, who will be responsible for the upgraded ELSNET website. Steven Krauwer is second from the right in the front row.

The ELSNET activities under the umbrella of the European Commission's Fifth Framework Programme have now started for real. We hope that the transition from the old to the new Framework Programme has been smooth, but it is certainly not our intention to make it completely invisible. You may have noticed that although we haven't even considered changing ELSNET's name, the expansion of the acronym is now 'European Network of Excellence in Human Language Technologies', instead of 'Language and Speech'. The reason for this is, of course, not that we have dropped language and speech and their integration from our agenda, but rather the acknowledgement that we can no longer afford just to concentrate on the two traditional information carriers in human-to-human and human-machine communication. Written and spoken language are more and more found as embedded parts of larger communication or information systems, integrated with other modalities. Obvious examples are the web, mobile telephony, and multimedia content, and I am pretty sure that by the time we are half way through this Framework Programme, new application areas will have appeared on the horizon.

ELSNET, as the community of key players in this field in Europe, will of course continuously be at the forefront of these new developments. As an organisation, with the specific objective of supporting and facilitating the work of this community, we will continue to act as a forum where researchers and developers can share their problems, their solutions, their needs, their expertise, their frustrations, and their visions.

To this end we shall continue to develop our website (www.elsnet.org), which – after a long period of very modest activity – is now up and running again. It is our ambition to make it the place where everybody in our field goes first when looking for information. We cannot do this on our own. We have already joined forces with the other support projects funded by the Human Language Technologies programme (HOPE /Euromap, ISLE/EAGLES and CLASS) in order to build a joint website, where each of the projects will provide you with specific types of information, but even that is not enough. We need more help – from you! ELSNET was not set up to be an organisation where a handful of people sitting in an office do all the work, whilst others just watch and wait for something interesting to emerge. If we really want to populate our website with (pointers to) relevant information, we need your help. I am sure that every single one of you is sitting on some information he or she would be happy to share with others, provided someone makes available the facilities to collect and publish the information, and this is exactly what ELSNET is offering you.

I hope that you will all read this as a personal invitation to contribute to ELSNET's information dissemination activities. What we ask from you is relatively little: share with us the information you have and which you think might be useful for others, and what you will get in exchange is access to the information contributed by all your ELSNET colleagues, which sounds like a very good return on your investment!

Spring 2000

elsnet

New Faces for ELSNET

Goodbye, Mariken: Welcome, Brigitte!

***Steven Krauwer**, ELSNET Coordinator*

In November 1997 we welcomed Mariken Broekhoven as the new ELSNET assistant coordinator, and now the time has come to say goodbye to her again.

When Mariken started working with me, I was a bit concerned about the fact that she had no prior knowledge of the field of language and speech technology, but to my suprise it didn't take her more than a few weeks to find out what ELSNET was about, what had to be done, and how we should organise ourselves in order to keep things running

Those of you who have dealt with her, will have discovered, as I have, that she was very quick at picking things up, very efficient and determined, and never scared to tackle and solve unknown problems. In brief: a great assistant to have in an organisation like ELSNET, where every day may be full of surprises. Thanks, Mariken, for making my life look as smooth as it cannot possibly have been in reality over the last two years!

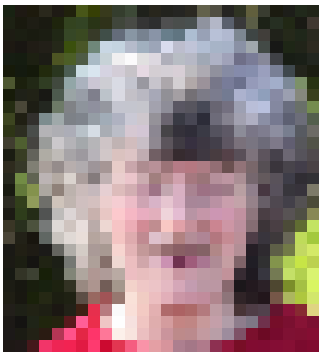
In her new job, Mariken will be heading the support team for the celebration of the 365th anniversary of our university here at Utrecht, and I would like to wish her (and myself, as part of this university) a successful birthday party!

The new assistant coordinator, Brigitte Burger, has already taken over Mariken's tasks and responsibilities. She graduated in linguistics, and has worked ever since as managing assistant for the Utrecht Institute of Linguistics OTS, and the Foundations for Language, and Speech, Technology (both with their headquarters in Utrecht). In the latter two functions she has had ample opportunity to get acquainted with the delights and frustrations of working in EU-funded projects, and even that didn't frighten her from taking this job.

Courageous and efficient – I'm looking forward to collaborating with her!

And Finally ... A New Editor for *ELSNews*

***Geoffrey Sampson**, University of Sussex*



Jenny Norris

The transition to Framework Programme V coincides with a change of editor for *ELSNews* and a change of production site to the University of Sussex. We offer a warm welcome to the new *ELSNews* Editor, Dr Jennifer Norris. Jenny occupied the editorial chair in March, shortly after completing a PhD on the automatic generation of 'aboutness expressions', at the neighbouring University of Brighton.

Jenny will be working at Sussex University under my direction. For both of us, getting to grips with the mysteries of desktop publishing in recent weeks has been a novel learning experience! We request readers' forgiveness in advance in case any rough edges show up in this first issue.

Moving to the University of Sussex represented a return to familiar territory for Jenny: she took her second bachelor's degree, in Linguistics and Cognitive Science, at Sussex University in 1989 (having already taken a degree in Genetics and Microbiology at University College London several years earlier).

In the mean time, Jenny has had a varied career, which has ranged from Microbiology to English language teaching – both overseas, in Italy and Spain, and with classes of immigrants here in Britain – to running her own business, designing and making cycling wear.

Cycling is one of Jenny's many leisure-time interests. She also enjoys sailing, and is putting a lot of effort at present into restoring the old farmhouse into which she and her family moved recently. The latest addition to the family, Erina, was born just six months ago.

Welcome to the team, Jenny! We look forward to *ELSNews* going from strength to strength over the coming months.

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

Editor: Jenny Norris.

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

ISSN 1350-990X

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

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

Material for the next issue is due:
15 July 2000

Multimedia Interaction for the New Millenium – Part II

Mark Maybury, *The MITRE Corporation*

Mark Maybury

Mark Maybury gave an ELSNET supported keynote speech at Eurospeech 1999 in Budapest. *ELSNews* published the first part of his abstract, based on his speech, in the last issue (8.4, December 1999). Here, as promised, is the second part.

Multimedia Information on Demand

Information on demand, the ability to provide information tailored to specific user needs, promises new capabilities for research, education and training, and electronic commerce (e.g, on-line information access, question answering, and customer service). Whereas significant commercial activity has focused on providing access to documents, web pages, and structured data sources, less attention has been given to multimedia information on demand. To achieve effective multimedia information on demand, however, requires a confluence of capabilities from several fields including image, speech and language processing, information retrieval, information extraction, translation, summarisation, and presentation design.

Scientists have begun creating systems to provide tailored, content-based access to multimedia including text, imagery, audio, and video (Maybury 1997). For example, by synergistically combining techniques for processing speech, language, and imagery, MITRE developed a sophisticated news understanding system, the Broadcast News Navigator (BNN) (Maybury et al. 1997). The web-based BNN gives the user the ability to browse, query (using free text or named entities), and view stories or their multimedia summaries. For example, Figure 2 displays all stories about Diana on CNN Prime News during August–September 1997). For each story, the user can view its closed caption text, named entities (i.e., people, places, organisations, time, money), an extracted multimedia summary, or view the full original video of a story.

Figure 2. Tailored Multimedia News

The user can also graph trends of named entities in the news for given sources and time periods. For example, Figure 3 graphs the onset and abatement of stories on Princess Diana and Mother Teresa, 8–15 September 1997.

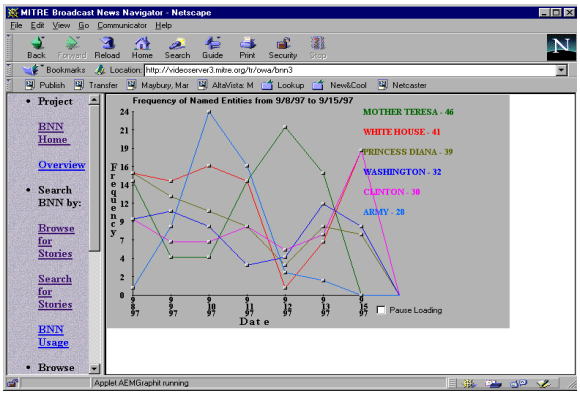


Figure 3. Temporal Visualisation of Named Entities

Analysing audio, video, and text streams from digitised video, BNN segments stories, extracts named entities, summarises stories, and designs presentations to provide the end user with content-based, personalised web access to the news (Maybury et al. 1997). For example, within the video stream colour histograms are used to classify frames and detect scene changes. In

the audio stream, algorithms detect silence, speaker changes, and transcribe the spoken language. Finally, the closed caption stream and/or speech transcription is processed to extract named entities. This fully automated broadcast news system stands in contrast to the current method of manual transcription and summarisation of broadcast news (e.g., via closed captioning services) which is expensive, error prone, and can result in dissemination delays. BNN has been integrated into a larger system called GeoNODE (Hyland et al 1999), which correlates named entities across stories to create story clusters and then partitions these clusters from a constructed hypergraph to identify topics automatically.

Our analyses show that in key tasks such as segmenting stories, audio and imagery processing can enhance algorithms which are based only on linguistic cues (e.g., explicitly stated anchor welcomes, anchor to reporter handoffs, story introductions). For example, silence detection, speaker change detection, and key frame detection (e.g., black frames, logos) can improve the performance of text-only story segmenters. By modelling story transitions using hidden Markov models and learning the most effective combination of cross media cues (Boykin and Merlino 1999), successive versions of the system have incrementally increased performance.

Given properly delineated story boundaries, BNN is able to summarise the story in a variety of fashions. This includes extracting (Aberdeen et al. 1995) the most significant named entities, extracting visual elements (e.g., key frames) and/or summarising the story text and creating from these elements a multimedia summary. We have integrated Carnegie Mellon University's SPHINX-II speech system into BNN and have begun experiments in extracting named entities from transcribed stories.

Figure 4. Manual Transcription and Extraction

For example, Figure 4 shows a manual transcript of a segment of a story with human markup of locations (bold italic) and organisations (bold underlined). In contrast, Figure 5 shows an automated transcription of a news segment followed by automated markup of locations (bold italic) and organisations (bold underlined). Notice the errors in the automated transcript of omission of punctuation and content (e.g., "A", "AND") and substitutions (e.g., "EVELYN" for "CONSIDERING", "TRIAL" for "PANEL"). Also note errors in the named entity identification in Figure 5 ("BLACK" is identified as a location, "UNITED STATES" and "CONGRESSIONAL BLACK CAUCUS" are missed). In the DARPA HUB-4 information extraction evaluation, the version of the system described in (Palmer et al. 1999) achieved 71–81% accuracy on broadcast news speech data with a wide range of word error rates (WERs) ranging from 13% to 28%. The model achieved 88% accuracy on reference transcriptions with WER 0% (like Figure 4).

Figure 5. Automated Transcription and Extraction

Even when dealing with closed-captioned text, we face a 10–15% word error rate because of errors introduced during manual transcription. The word error rates for the best automated speech transcription systems (depending upon processing speed) range widely from 13–28% on studio quality speech (e.g, anchor segments) to 40% or higher on shots with degraded audio (e.g., reporters in the field, speakers calling in over the phone, music in the background). Furthermore, neither closed captions nor speech transcripts have case information to use, for example, to recognise proper nouns. In addition, speech transcripts contain no punctuation and can contain dysfluencies (e.g., hesitations, false starts), which further reduces performance levels. However, it is important to point out that the type of errors introduced during human transcription are distinct from those made by automated transcription, which can have different consequences for subsequent named entity processing. In the HUB-4 evaluations (DARPA 1999), named entity extraction on clean, caseless, and punctuationless manual transcripts was approximately 90%, in contrast to the best extraction performance on newswire with case, which was approximately 94%. >

MITRE's named entity extraction system, called Alembic[1] (Aberdeen et al. 1995), consists of rule sequences that are automatically induced from an annotated corpus using error-based transformational learning. This significantly enhances cross domain portability and system build time (from years or months to weeks) and also results in more perspicuous rule sets. Our Alembic group is presently pursing the induction of relational structure with the intent of automated event template filling.

We have performed a number of user studies (Merlino and Maybury 1999) to discover the optimal mix of summary elements (e.g. key frames, key extracted named entities, text extracts). Compared to a digital VCR, BNN enables users to perform their information seeking tasks more accurately, faster (in some cases six times as fast), and with twice the satisfaction compared to traditional displays.

Conclusion

Multimedia interfaces and multimedia information access continue to advance. Our ability to harness multimedia for user benefit will become an increasingly important challenge as we move into the next millennium.

Acknowledgements

I would like to thank Lynette Hirschman, David Palmer, John Burger, and Andy Merlino for providing the spoken language processing examples for BNN. Sam Bayer and Laurie Damianos are responsible for the multimodal logger. I also thank Stanley Boykin and Andy Merlino for providing BNN story segmentation performance results.

REFERENCES

Aberdeen, A., Burger, J., Day, D., Hirschman, L., Robinson, P. & Vilain, M. 1995. MITRE: Description of the Alembic System as Used for MUC-6. In *Proceedings of the Sixth Message Understanding Conference* (MUC-6), Columbia, MD, 6–8 November 1995, 141–155.

Boykin, S. and Merlino, A. Improving Broadcast News Segmentation Processing. IEEE International Conference on Multimedia and Computing Systems. Florence, Italy. 7–11 June 1999.

DARPA Broadcast News Workshop, 28 February–3 March 1999, Herndon, VA.

Hyland, R., Clifton, C., and Holland, R. 1999. Geonode: Visualizing News in Geospatial Context. AFCEA Federal Data Mining Symposium. Washington, D.C

Maybury, M. T. (ed.) 1997. Intelligent Multimedia Information Retrieval. Menlo Park: AAAI/MIT Press. (http://www.aaai.org/Press/Books/Maybury-2)

Maybury, M., Merlino, A., and Morey, D. 1997. Broadcast News Navigation using Story Segments, ACM International Multimedia Conference, Seattle, WA, November 8–14, 381–391.

Merlino, A. and Maybury, M. 1999. An Empirical Study of the Optimal Presentation of Multimedia Summaries of Broadcast News. In Mani, I. and Maybury, M. (eds) *Automated Text Summarisation*. Cambridge, MA: MIT Press, pp. 391–401.

Palmer, D., Burger, J. and Ostendorf, M. 1999. Information Extraction from Broadcast News Speech Data. DARPA Broadcast News Workshop, 28 February– 3 March 1999, Herndon, VA.

Endnote: This document is a summary of Maybury, M. T. 1999. Keynote. Intelligent Multimedia for the New Millennium. Proceedings of Eurospeech '99. Budapest, September 6–9, 1999. vol 1. p. KN1–15

[1]The Alembic Workbench, which enables graphical corpus annotation, can be downloaded from http://www.mitre.org/resources/centers/it/g063/nl-index.html.

FOR INFORMATION

Mark Maybury is Executive Director of MITRE's Information Technology Division. He is responsible for the direction of MITRE advanced research and development for intelligence and defence systems, supporting a broad range of sponsors, including the US Air Force, Army, Navy, DARPA and classified agencies.

Email: maybury@mitre.org

Web: http://www.mitre.org/resources/centers/it

Report

Computers, Language and Speech

Report of the Royal Society/British Academy joint discussion meeting, 22–23 September 1999

The event was organised by Roger Needham, Gerald Gazdar, and Karen Spärck Jones
This report was contributed by ***Gerald Gazdar***

Gerald Gazdar

As readers of *ELSNews* know well, there has been recent rapid development in the use of statistical techniques in both written and spoken language processing. In the light of this development, there are important issues to address about the interaction between formal symbolic theories of language and the new statistical approaches. The central question addressed in the meeting was how best to combine rule-based and statistics-based approaches to natural language. More specifically, it provided an opportunity for the text and speech communities to exchange their respective findings and ideas in the light of

- the growth of corpus-based strategies in text interpretation and generation, until quite recently primarily symbolic and rule-based;
- the interest in enriching speech processing, hitherto predominantly statistics-based, with prior knowledge of a symbolic kind.

The meeting was timely given the increasing demand for practical NLP systems able to cope with bulky, changing or untidy material, and the rapid growth of machine resources able to support the demanding data analysis and rule application that this development implies.

The speakers were Fernando Pereira (AT&T), Julie Caron-Berndsen (UCD), Stephen Pulman (Cambridge), Harald Baayen (Nijmegen), Stephen Renals (Sheffield), Roni Rosenfeld (CMU), Mari Ostendorf (Boston), Geoffrey Sampson (Sussex), Hiyan Alshawi (AT&T), Jon Oberlander (Edinburgh), Stephen Young (Cambridge), Paul Taylor (Edinburgh) and Kathleen McKeown (Columbia).

Some of the papers illustrated the interaction between statistical data and model rules for speech processing, whether in recognition or synthesis. Others were concerned with text or transcribed speech. At the same time, the papers addressed many different language levels from the components of words, through intermediate units like phrases or sentences, to discourse units like whole dialogue turns, to extended text, and even to the real world domains that underlie linguistic expressions.

Some of the papers started from the use of statistical data and pushed this past words to capture larger unit regularities and hence higher-level language structure; others also started from the data but attempted to leverage pattern capture by exploiting independent linguistic features, constraints or rules. But the complementary strategy, starting from the rule end but modifying and developing an initial model in the light of observed usage, was also represented.

The papers presented illustrated a wide range of techniques for capturing statistical regularities and for representing syntagmatic and paradigmatic language structure, in a way suited to linking data and rules, whether working up from the former or top down from the latter. Again, just as the papers attacked different language levels, they also addressed different subtasks within the scope of a comprehensive language processing system, for instance, from word recognition in interpretation to style constraints in text generation. They also illustrated the role of statistically-motivated approaches for some application tasks, like translation.

Memory-based techniques made several appearances: two of the papers employed them for prosodically coherent speech synthesis whilst a third proposed a system for morphological interpretation in which a parser and a parse memory compete to deliver the most plausible word structure. Memory fails, by definition, for items previously unseen. The 'unknown word' problem is pervasive in NLP and was explicitly addressed by several speakers. >

Finally, there were papers that addressed the inputs and outputs for work in this whole area, namely the general requirements for systematically described corpus data as input, and the evaluation of the results of data analysis, both from a methodological point of view and as illustrations of the performance that language processors exploiting statistical resources can currently achieve.

There were, moreover, some significant common threads in the meeting papers. The first was an emphasis on the central role of the word. As Pereira pointed out, the widespread adoption of the word as the central unit of analysis makes it easier to anchor theory in observation. The second thread was a willingness to rewire the canonical circuit diagram for NLP systems, by relaxing accepted divisions of level and unit. This was illustrated most dramatically in the papers given by Alshawi and Young. And the third was a liberation of text-based NLP brought about by the marginalisation of the notion of well-formedness. Instead of simply rejecting an approach *a priori* on the grounds that it could be proved to fail in certain circumstances, researchers now embrace techniques that have the potential to work well most of the time. Success has become a matter of degree.

It is of course difficult to say whether a meeting only a few months ago, in an area in which media-friendly 'discoveries' and 'breakthroughs' do not figure and where the papers have not yet been published, has had an effect. But the discussion that followed each paper was extensive and invariably constructive. And the feedback from those who attended was uniformly positive: one participant commented that it was the only such event at which he had felt moved to attend every paper and the only one at which he had no cause to regret any of his attendance decisions. More generally, the time was clearly right for such a meeting and the comments made by participants suggest that it will encourage relevant research, both by those who were there and and by those who seek out and read the published papers.

The proceedings of the meeting (including reports of the discussion) are published in April this year in the *Philosophical Transactions of the Royal Society, Series A, Volume 358, Issue 1769* [this is the world's oldest scientific journal! – Ed.]. A contents list and an order form for the volume is available at http://www.cogs.susx.ac.uk/lab/nlp/colasp/colasp.html or from Jacqueline Knapp at the Royal Society. There is a significant discount for members of the ACL.

FOR INFORMATION

Prof. **Roger Needham** FREng, FRS is Pro-Vice-Chancellor of the University of Cambridge, and Managing Director of Microsoft Research, Cambridge. From 1980 to 1995 he was Director of the University of Cambridge Computer Laboratory.

Gerald Gazdar FBA is Professor of Computational Linguistics at the University of Sussex.
Email: geraldg@cogs.susx.ac.uk

Karen Spärck Jones FBA is Professor of Computers and Information at the University of Cambridge. In 1994 she was President of the Association for Computational Linguistics

Call for Letters

Feature

Does ELSNET No Longer Go East?

George Chikoidze, *Tbilisi State University, Georgia*

Environs of Mtskheta, the ancient capital of Georgia (2nd Symposium, 1997). In the foreground are Georgian couple Nino Amiridze and Temuri Kutsia. Both are currently working abroad: Nino at Utrecht University, Temuri in Linz.

During her 1993 visit to Great Britain, the future member of the Georgian Parliament, Nani Chanishvili, happened to mention that in Georgia, despite particularly hard times, some research groups were managing to continue their work in computational linguistics, speech recognition and logic. In response, she was invited to the Centre for Cognitive Science at the University of Edinburgh, where she described the state of affairs in Georgia, both in general, and in the specific above-mentioned fields. She also met researchers at the Centre, including Jonathan Ginzburg and Enric Vallduví.

Between them, Ginzburg and Vallduví initiated a course of action which laid the foundations for the present Centre for Language, Logic and Speech at Tbilisi State University. In 1994, mainly on their initiative, the Tbilisi Spring School in Language, Logic and Computation was held. In addition to Ginzburg and Vallduví, Gregor Erbach and Tim Fernando also gave lectures. The success of this forum suggested the need for a Centre to unite the hitherto uncoordinated researchers in language technologies into one community. Ginzburg and Vallduví also thought that such a united centre would be better for attracting the attention of Western partners, and in this respect they turned out to be true prophets.

Before elaborating on this latter point, I will give a brief history of linguistic research in Georgia.

Georgia has a strong history in traditional linguistics and logic. One prominent linguist, Th. Gamkrelidze, founded the Department of Structural and Applied Linguistics at Tbilisi State University in the late '60s, where mathematics, some logic and rudimentary computational linguistics were taught. Over that decade I was directing a project on machine translation from Russian into Georgian, and was able to use my expertise to run several courses on machine translation.

The Institute of Control Systems at the Georgian Academy of Sciences, where I was then (and am now) employed, has played a prominent role in the development of research in Language and Speech Technologies in Georgia. From 1959 to 1999 two Departments of the Institute, of Speech and of Language, have worked either separately or jointly on automatic translation, speech dialogue, language learning, speaker identification and more.

Within the University, the Institutes of Applied Mathematics, Oriental Studies, Cybernetics, and Linguistics have all been conducting research in various topics related to computational linguistics. For example, a group at the Institute of Linguistics has been working intensively over the last decade on computer modelling of the notoriously complex Georgian morphology.

I must admit (though such declarations are now drastically out of vogue) that all this research was rather generously financed by the USSR institutions. Thus, the above-mentioned projects at the Institute of Control Systems were sponsored for 15 years (1977–1991) by the Ministries of Defence and Justice, by the Navy and several All-Union scientific programmes.

At that time, the general ideology and policy of the USSR government meant that all contacts with the West, including scientific ones, were strictly restricted and impeded, as was the flow of information from periodicals, books, participation in forums abroad and so on. As a result, I (and very probably not only myself!) attended the lectures of the Tbilisi Spring School which covered the modern issues in logic and linguistics merely out of curiosity, without having any serious understanding of their tenor.

It seems that this quite undeserved feeling of 'mental inferiority', and our eagerness to overcome it, were the most powerful incentives behind the organisation of the new Centre for Language, Logic and Speech. Crucially,

we saw this as a means of improving existing weak contacts with the West and establishing new ones. The Centre was founded in the autumn of 1994, and the Board included representatives of almost all the institutions which had been carrying out related research independently from one another in the fields of language, speech, logic and computation.

At the 3rd Symposium (1999) George Chikoidze discusses something very important with Olga Boguslavskaja (Moscow).

The first undertaking of the new Centre was an International Symposium on Language, Speech and Computation, held in the autumn of 1995 at Gudauri, a fashionable mountain/ski resort near the main ridge of the Caucasus. The beautiful scenery and comfortable hotel (with swimming pool and exquisite menu) were wonderful. Yet the real success was due to the participation of some prominent speakers (Robin Cooper, Peter Gardenfors, Aravind Joshi, and others). As a result, guests and hosts unanimously decided to make this a regular, biennial, event.

The First Symposium provided a great opportunity to make personal contacts with our guests from the West. However, the most important effect (perhaps partially influenced by having made these contacts) was the promotion of the Centre as a Coordinating Node in the 'ELSNET Goes East' (EGE) Project. It was fortunate for us that the EGE Project was running then, and even more lucky that it was coordinated by two business-like, active and kind young people: Eric Jan van der Linden and Ingrid van Loon. Ingrid, in particular, has subsequently been enormously helpful and supportive, even after the end of EGE.

The EGE Project provided us initially with essential equipment, a travel budget and conference registration fees, as well as a precious set of books and other essentials. However the central event supported by EGE was the Second Symposium on Language, Logic and Computation, held in 1997 in the capital of Georgia, Tbilisi, at the State University. It again attracted some prominent participants, such as Dick de Jongh, Barbara Partee, Ju. Apresjan and others. Serious scientific discussions were punctuated by some enjoyable and relaxing events, the most memorable of which was a visit to Kakheti, the main wine-producing region of Georgia.

Finally, the Third Tbilisi Symposium was organized in conjunction with the University of Amsterdam. It was held last year (1999) at the Black Sea resort of Chakvi, where the hotel cottages are spread out along the sea shore. Besides 'old-timers' (Robin Cooper, Paul Dekker, Dick de Jongh) there were some prominent new faces, such as Lauri Karttunen, Igor Mel´chuk, Annie Zaenen and others.

The Proceedings of the First and Second Symposia are published by Stanford University Press and Tbilisi State University respectively. The Fourth Symposium (2001) is to take place in Kutaisi, the second largest city (after Tbilisi) of Georgia. As for this year, we are working on preparations for the Summer School (covering the same topics as the Symposia), to be held between 29th August and 8th September [see 'Future Events', pg. 16 – Ed.], for which the Proceedings are being prepared.

Well then ... why this feeling of dissatisfaction? What is missing!?

What is missing, I would say, is the single most precious thing: the level of acquaintance with modern theories of language and logic. Even with the very friendly and helpful contacts we have made at our Symposia, we are left with a serious lack of any Western–Eastern common scientific project.

This problem is partly addressed through some of our students and young researchers visiting leading scientific centres of the West (mainly in the Netherlands, but also in Sweden and some other countries). Such visits are, however, few, and not always successful: particularly when the best of our researchers strive to stay abroad!

What we need is more books and introductory courses aimed at updating our knowledge, especially for the younger members of our community, who are our future. We have taken some first steps here: for example, the tutorials held during the Third Symposium; the acquisition of a set of books, presented by the University of Amsterdam, which, along with some we already had, make a good start for a future library. We have invited some visiting lecturers (A. Voronkov and M. Marx), who will teach various introductory courses; and, of course, there is the Summer School, which will run seven courses on linguistic and logic topics. >

Tbilisi State University, Georgia

Our hopes for the future include expanding our book collection into a specialist library; inviting more visiting lecturers; and organising a distance learning facility.

It is, unfortunately, difficult to realise such an ambitious programme in a developing country, where the salary of scientists amounts to just US$20–25 per month (paid irregularly if at all!), and where the number of academic staff has more than halved over the last decade. We must try our utmost to retain and develop the very valuable scientific potential we have shown, and to attract our talented young people, many of whom have already gone abroad or changed fields.

So, we are doomed to look for aid from abroad. Our experience here (sometimes rather bitter) suggests that a single organisation – perhaps a unique centre which sponsors related research and would offer advice, coordination and mediation – would be the best solution. Naturally, after EGE we had our hopes set on ELSNET and were delighted when Tbilisi State University became an ELSNET node in 1998.

However … I will finish boldly with an extract from the final Progress Report submitted to the EGE Project:

"Obviously, all the aforesaid proves beyond any doubt the high quality and fruitfulness of EGE Project activity for Georgia and other C&EE countries. Our only concern in this regard is that its ending is untimely from many points of view. Thanks to EGE activity many auspicious processes have started: nevertheless, their natural and full development requires continuation. The validity of such anxiety may be confirmed, e.g., by the state of affairs regarding common projects and corresponding perspectives.

It seems that this problem could be solved by several different, though not mutually exclusive, approaches. As a more global approach, I see two possibilities create some immediate successor of the EGE Project, or admit the 'orphaned' Coordinating Nodes of EGE into the 'parent' ELSNET, in which case the latter should shoulder a great deal of EGE past activities.

In my opinion, at least one of these approaches should be adopted: otherwise, the most beneficial fruits of EGE activity will go to waste. We shall stumble over the threshold of the Information Age and then retreat in disorder."

This was written nearly three years ago. A voice in the wilderness?!

FOR INFORMATION

George Chikoidze is Head of the Department of Language Modelling at the Institute of Control Systems, Georgian Academy of Sciences. He is the contact person at the Tbilisi State University for ELSNET.

Email: chiko@contsys.acnet.ge,
Tel: +995 32 322136
Fax: +995 32 942391
Web: http://www.geo.net.ge/llc99

(continued from page 11)

YW *(cont)*: If we could understand that better we might make a lot of this argument clearer and enable the main groups to have a better view of how they should cooperate: we all seem more muddled here on this – why is that?

MB: The Commission's goal is to advance knowledge, employment and their own political position. Capitalistic advances take place anyhow. Look at the dictation market. IBM, L&H, Dragon and Philips have sold millions of dictation systems. Do you know anybody who uses them? This is a perfect market: sell things that nobody uses. I have toyed with the idea of selling a system to translate the barking of a dog into English (or Dutch – I don't care).

YW: Nice example – do you remember the old Godard film *Alphaville*? There was a machine that had a sign 'Insert one franc' and when you did it said 'Merci'. What a product!!

Seriously – have there been any demonstrations that any EU programme creates any employment? I am not asking cynically – just remembering all the demonstrations in the papers that gross differences in job creation (i.e., between the EU and the US) depend on major factors that have nothing at all to do with science and research: chiefly restrictions on labour mobility. Why not let us all be honest about where the problem is and not pretend it is in scientific/industrial creativity?!

MB: In Belgium, a large industrial entity for language processing generates employment (The Flanders valley with Lernout and Hauspie) but relatively little research. In Holland, with the same language, I see many more researchers but no real industrial base. This is only an

Opinion Column

What are we doing all this for?

A dialogue on Language Engineering

Yorick Wilks *and* ***Marc Blasband***

MB: Let me start with what seems to be an obvious fact and one enshrined in Framework Programmes: the end user must be central in fact and not just in words.

YW: The end user is clearly central for certain kinds of funders (i.e., EU as opposed to national science funds in different states) but the end user is never the main motivation for a researcher – and you can't change that. All you can do is give up talking to researchers if you don't accept it. That may be all right if you can find enough talented development-oriented workers who don't really care about research, but can take in enough existing research to get going and productise. This is one version of the Japanese model and that seems to be what you want, but it has never till now been the EU or US model.

MB: Let me answer these two points together. If we do artificial intelligence research to play chess or to support a general or captain of industry, then I agree with those researchers. But if we do AI for the masses, we have to know what people want, how they react and how they evolve. I am not sure if this is a problem related to a model or to the issue of focusing the researchers' attention on that domain of research.

Another point is that the Commission wants to spend quite a portion of its budgets for user applications, but they talk very little to users, which seems very dangerous to me.

YW: Surely they don't mean to – its just incompetence if that happens – since the whole programme talks *ad nauseam* about users?

MB: Look at any meeting organised by the Commission. There are no users. The same is true of the Dutch and Flemish programmes. I have talked at length with Commission officials on that, and their explanations are that users are difficult to get into the research loop (that is true) and they rely on the suppliers to represent the wishes of their users through the market mechanisms (this is utopian). I think the end-user should profit from the research through exploitation.

YW: What *is* scientific exploitation for you? Surely for the scientist it is more funding and more results? Also, don't forget Government exploitation of research – quite different from commercial – DARPA, atom bombs, intelligence, the police etc. I am sure all EU governments (and the US) are also saying that we don't get enough commercial exploitation from science, which is the purpose of the Framework programmes. But the more commercial they get, the less do we find good scientists wanting to play at all. Many of my smartest colleagues in places like Cambridge will not touch an EU grant for that reason; exploitation for them could only mean their own company. And I cannot see what you mean by end-users benefiting from research by exploitation: end-users benefit as consumers, passive beneficiaries from someone else's exploitation!

MB: Everyone has their own definition of exploitation. The Commission wants something they can use politically, thus new industries, more employment. The governments want more atom bombs etc.; the researchers, more knowledge (or more glory?); the industry, more saleable products; the consumer, more fun.

YW: I agree that the participant groups have very different goals. The interesting thing to me is why it is that other capitalist societies like the US are more successful in some ways, e.g., small company formation, job creation etc. The same contrasting 'interests' there have their compatibilities fostered in quite different ways – for example, we all know that it is simply easier in the US to raise capital and get a technical product to market than in the EU. Should we be asking more (within the industry/research/Commission communities) why that is?

(continued on page 10)

FOR INFORMATION

Yorick Wilks is Professor in the Department of Computer Science, University of Sheffield.

Email: yorick@dcs.sheffield.ac.uk

Web: http//www.dcs.shef.ac.uk/~yorick/

Marc Blasband is a software consultant with an international carrier. Over the last four years he has provided services to the Dutch Railways for their train schedule spoken dialogue systems. He is mainly interested in the usage of the technology.

Email: cplr@worldonline.nl

Project Update

DISC – spoken language DIalogue Systems and Components

An Overview of Results

Niels Ole Bernsen *and* ***Laila Dybkjær****, University of Southern Denmark*

Spoken language dialogue systems (SLDSs) are becoming mainstream technologies which provide a wealth of different services in an increasing number of languages. In general, SLDSs are complex systems which incorporate speech recognition, speech generation, natural language understanding and generation, dialogue management and database components. The building of successful applications not only demands expertise in those technologies but also in systems integration, human factors, design and development support tools and, increasingly, call-centre technology. Difficult research problems remain to be solved in most of the areas mentioned above, and best practice guidance is notoriously absent in the field of dialogue engineering.

The DISC project – spoken language DIalogue Systems and Components, best practice in development and evaluation– was launched in June 1997 with the aim of developing a first Best Practice Guide on how to develop and evaluate SLDSs and their components (cf. *ELSNews* 6.4,1997, pp. 6-7). Ending in December 1998, DISC was immediately continued in DISC-2 which ran until the end of 1999.

The first phase of DISC was dedicated to the development of current practice reviews of the DISC SLDS aspects (see below), a detailed best practice dialogue engineering development and evaluation methodology, and a range of design support concepts and software tools. The second phase, i.e. DISC-2, focused on testing the validity and usability of the draft Best Practice Guide, the concepts and the tools, and on the integration, packaging and dissemination of the final DISC Best Practice Guide. Throughout DISC, the Advisory Panel, which by the end of 1999 counted more than 50 researchers and industrial developers from across the world, has been an invaluable source of critique and comments on progress.

The DISC current practice reviews charted current SLDS development and evaluation practice, producing about 50 in-depth analyses of existing SLDSs and components together with the following approach to dialogue engineering best practice. An SLDS is viewed as having six major *aspects:* speech recognition, speech generation, natural language understanding and generation, dialogue management, human factors, and systems integration.

Each aspect can be analysed in terms of a *grid.* A grid contains an aspect-specific description of the state-of-the-art technical problem space facing the developer, including technical properties, interrelationships among properties, and advice on which properties to include in particular applications. Within the grid problem space – or outside it, since new ideas appear all the time – the developer must make the decisions most appropriate for the application to be developed. In DISC, the grid problem space is structured in terms of the *issues* facing the developer, the *options* the developer must choose from per issue, and the *pros and cons* with respect to each option.

Orthogonal to the 'static' grid description, each aspect may be analysed in terms of a development *life-cycle* which decomposes the development process into iterative phases and issues to be addressed in each phase. Integral to the life-cycle is the continuous *evaluation* of progress and results. As DISC progressed, evaluation of SLDS aspects gained prominence due to the many unsolved research issues in this field. In response, DISC has developed a generic *evaluation template* which can be used to characterise each evaluation criterion for use in evaluating aspect-specific properties of SLDSs. Furthermore, the grid analyses have been used to generate systematically a set of evaluation criteria for each aspect.

In addition to the above, best practice guidance must incorporate guidance on available platforms, methods and supporting tools for each aspect. These have been surveyed in a series of DISC reports. Moreover, DISC has itself produced a series of development support

tools and guidelines. These include

- guidelines and testing protocols for the development of speech recognition components for SLDSs;
- a software tool for evaluating speech synthesis components in SLDSs;
- guidelines for the acquisition of lexical data for SLDSs;
- CODIAL, a software tool in support of cooperative system dialogue design;
- SMALTO, a software tool in support of speech functionality decisions (pertaining to what the speech modality is or is not good for), used in early design.

The core result of DISC is the web-based DISC Best Practice Guide (www.disc2.dk), which resulted from turning everything mentioned above into a comprehensive website. In addition to the ingredients described above, the DISC Best Practice Guide includes a comprehensive *glossary* of dialogue engineering terminology, *references* to the literature including all DISC publications, and brief *checklists* for each aspect.

Due to generous support from ELSNET, the DISC website will continue to be updated in the future. The DISC consortium, therefore, welcomes all comments from those who have tried out any part of the DISC Best Practice Guide. You may email the authors of the present article who will make sure that your comments reach the right people. However, the DISC website offers several other possibilities of interacting with the DISC consortium which includes NIS (Odense, Denmark, coordinating partner), CNRS-LIMSI (Paris, France), IMS (Stuttgart, Germany), KTH (Stockholm, Sweden), Vocalis Ltd (Cambridge, UK), Daimler-Chrysler AG (Ulm, Germany) and ELSNET (Utrecht, The Netherlands).

FOR INFORMATION

Niels Ole Bernsen and Laila Dybkjær are at the Natural Interactive Systems Laboratory (NIS) at the University of Southern Denmark, Odense, Denmark. Niels is a member of the ELSNET Executive Board.

Email: nob@nis.sdu.dk and ld@nis.sdu.dk

DISC Best Practice Guide website:
http://www.disc2.dk

Announcement

WEB–SLS

The European Student Journal of Language and Speech

Nikos Fakotakis, *University of Patras, Greece*

WEB–SLS is a web-based Journal on Speech and Language Engineering, devoted to students who wish to publish their work and those who wish to learn from it. It accommodates two types of papers

- regular papers, which pass through a fast review process involving at least two reviewers;
- extended PhD summaries, formatted as regular papers with references to all relevant authors' publications, accepted upon supervisor confirmation only.

The Journal is sponsored by ISCA, EACL and ELSNET, and has training as its main objective. It is open to students up to PhD level, and we especially encourage graduate students and students doing postgraduate master courses to submit manuscripts, since this work rarely reaches the open literature. We hope to get many first papers from able students all over the world.

Given the wide subject coverage of the Journal, we have specified nine very broad areas:

- speech recognition and synthesis (SRS)
- natural language processing (NLP)
- speaker and language recognition (SLR)
- information extraction and retrieval (IER)
- dialogue modelling and control (DMC)
- models of perception and production (MPP)
- resources of speech and language (RSL)
- system assessment (SAS)
- other (OTH)

WEB–SLS is published continuously and invites contributions from students across Europe whose work lies in the above fields. Please submit either your paper, or your extended PhD summary (as detailed above), to the Editor in Chief. Details, instructions and guidelines for authors are provided on our website (http://web-sls.essex.ac.uk/web-sls).

Papers are published in HTML format. They are grouped in volumes according to the year of publication, and are freely accessed on the web. The electronic nature of the medium facilitates the inclusion of colour, sound, video and demonstrations. The evaluation process is fully electronic and therefore fairly fast.

The Journal is organised into three committees: an Advisory Committee, an Editorial Board and a Student Committee. The small Advisory Committee (nominated by the three sponsoring institutions) oversees the continuity and development of WEB–SLS as a proper scientific journal. It advises the Editor in Chief and the Technical Publishing Editor in all matters relating to the policy of the Journal, especially those concerning training.

The Editorial Board is responsible for maintaining academic standards through the reviewing process. The Board currently includes over thirty leading researchers drawn from the European Language and Speech communities who are also actively involved with students. Editorial Board members do the reviewing as well as the lobbying for papers, and search for extra reviewers. They also fulfill a training function, and manage a carefully controlled section which offers constructive and useful feedback to authors. This comments section is reviewed by at least one person and is published only with the permission of the original author(s) of the paper.

Finally, the Student Committee aims at actively involving more advanced students and helping them learn about the reviewing process. This Committee is a pool of participants in the reviewing procedure and identifies the best papers of the year. It has twenty members, appointed annually from 1st September, and is renewed every year by a special annual call.

Call for Student Members

Students who wish to participate in the Committee should send an application email along with a short CV to sgarbas@wcl.ee.upatras.gr by 31st July. Previous members of the WEB–SLS Student Committee are still eligible for the next year, as long as they retain their student status, but they must re-apply for each year they wish to serve on the committee.

FOR INFORMATION

WEB-SLS Editor in Chief: Nikos Fakotakis
University of Patras, Greece
Email: (fakotaki@wcl.ee.upatras.gr)

Technical & Publishing Editor: Mark Tatham
University of Essex, Colchester, UK

Web: http://web-sls.essex.ac.uk/web-sls

Future Events in 2000

June 12-16 *International Natural Language Generation Conference (INLG 2000)*, Mitzpe Ramon, Israel.
Workshops, June 12; conference, June 13-16.
Email: nlg2000@cs.bgu.ac.il; URL: http://www.cs.bgu.ac.il/~nlg2000/

June 15-17 *4th Workshop on the Semantics and Pragmatics of Dialogue (Götalog 2000)*, Göthenburg University, Sweden.
Email: gotalog@cogsci.ed.ac.uk; URL: http://www.ling.gu.se/gotalog/

June 23 *Workshop on Spoken Dialogue Systems, (Language and Speech Technology Programme)*, University of Groningen, The Netherlands.
Email: dijkstra@nwo.nl or gerdes@nwo.nl; URL: http://www.grid.let.rug.nl

July 3-5 *3rd Workshop on Human-Computer Conversation*, Bellagio, Italy.
Email: yorick@dcs.shef.ac.uk
URL: http://www.dcs.shef.ac.uk/research/units/ilash/Meetings/bellagio2000/index.html

July 3-7 *Workshop on the Nature of Speech Perception*, Utrecht, The Netherlands.
Email: bert.schouten@let.uu.nl
URL: http://www.let.uu.nl/%7Ebert.schouten/personal/workshop.htm

July 15-30 *8th European Summer School on Language and Speech Communication*: Text and Speech Triggered Information Access (TeSTIA). Chios Island, Greece.
Email: elsnet@let.uu.nl; URL: http://www.ilsp.gr/testia/testia2000.html

July 29-Aug 6 *18th International Conference on Computational Linguistics (COLING 2000)*, Saarbrüecken (conference, July 31-Aug 4); Nancy (tutorials, July 29-30); Luxembourg (workshops, Aug 5-6).
Email: org@coling.org; URL: http://www.coling.org

Aug 6-18 *12th European Summer School in Logic, Language and Information (ESSLLI-2000)*, Birmingham, UK.
Email: franconi@cs.man.ac.uk; URL: http://www.folli.uva.nl/Esslli/2000/esslli-2000.html

Aug 8-12 *9th EURALEX International Congress*, Stuttgart, Germany.
Email: elx2000@ims.uni-stuttgart.de; URL: http://www.ims.uni-stuttgart.de/euralex

Aug 29–Sept 8 *Tbilisi Summer School in Language, Logic and Computation*, Tbilisi, Georgia.
Email: chiko@contsys.acnet.ge; URL: http://www.geo.net.ge/llc99

Sept 5-7 *ISCA Workshop on Speech and Emotion*, Belfast, Northern Ireland.
Email: e.douglas-cowie@qub.ac.uk; URL: http://www.qub.ac.uk/en/isca/index.htm

Sept 13-16 *3rd International Workshop on Text, Speech and Dialogue (TSD 2000)*, Brno, Czech Republic
Email: tsd2000@fi.muni.cz; URL: http://www.fi.muni.cz/tsd2000/

Sept 18-20 *ISCA ITRW International Workshop on Automatic Speech Recognition (ASR2000)*, Paris, France.
Email: asr2000@limsi.fr; URL: http://www-tlp.limsi.fr/asr2000

Sept 20-23 *Architectures and Mechanisms for Language Processing (AMLAP 2000)*, Leiden, The Netherlands.
Email: AMLaP@fsw.leidenuniv.nl; URL: http://www.amlap.org

Sept 22-24 *5th TELRI Seminar on Corpus Linguistics*: How to Extract Meaning from Corpora. Ljubljana, Slovenia.
Email: telri-admin@ids-mannheim.de; URL: http://www.telri.de and http://nl.ijs.si/telri00/

Sept 25-28 *International Workshop on Speech and Computers (SPECOM 2000)*, St. Petersburg, Russia.
Email: specom@mail.iias.spb.ru; URL: http://www.spiiras.nw.ru/speech

Sept 26-28 *16th Conference of the Spanish Society for Natural Language Processing (SEPLN 2000)*, Vigo, Spain.
Email: sepln-secret@ei.uvigo.es; URL: http://www.coleweb.dc.fi.udc.es/sepln2000/

Oct 2-5 *Workshop on Speech Recognition and Synthesis (Prosody 2000)*, Krakow, Poland.
Email: gibbon@spectrum.uni-bielefeld.de; URL: http://www.ptfon.wmid.amu.edu.pl

Nov 16-17 *22nd Conference on Translating and the Computer*, London, UK.
Email: nicole.adamides@aslib.co.uk; URL: http://www.aslib.co.uk

Nov 20-22 *Machine Translation and Multilingual Applications in the New Millennium (T 2000)*, Exeter, UK.
Email: MT2000-request@rwsh.dircon.co.uk; URL: http://www.bcs.org.uk/siggroup/sg37.htm

Nov 23-25 *26th Annual Conference on Language Technologies:* Cologne, Germany.
Email: klaus.schmitz@fh-koeln.de; URL: http://www.fbi.fh-koeln.de/DEUTERM/ivsw2000E.htm

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

ELSNET Participants

Academic Sites

AT	Austrian Research Institute for Artificial Intelligence (OFAI)
AT	University of Vienna
AT	Vienna University of Technology
BE	Leuven University
BE	University of Antwerp - UIA
BG	Academy of Sciences Institute of Mathematics
BY	Belorussian Academy of Sciences
CH	Istituto Dalle Molle (IDSIA)
CH	University of Geneva
CZ	Charles University
DE	Christian-Albrechts University, Kiel
DE	German Research Center for Artificial Intelligence (DFKI)
DE	Institute of Applied Information Science (IAI)
DE	Ruhr-Universitaet Bochum
DE	Universitaet Erlangen
DE	Universitaet Hamburg
DE	Universitaet Stuttgart
DE	Universitaet des Saarlandes
DE	Universitaet des Saarlandes
DK	Aalborg University
DK	Center for Sprogteknologi
DK	Odense University
ES	Polytechnic University of Catalonia
ES	Universidad Politecnica Madrid
ES	Universidad Politécnica de Valencia
ES	Universitat Autonoma de Barcelona
ES	University of Granada
FR	CRIN
FR	IRISA/ENSSAT
FR	Inst. National Polytechnique de Grenoble
FR	Institute de Phonétique, CNRS
FR	LIMSI/CNRS
FR	Université Paul Sabatier (Toulouse III)
GE	Tbilisi State University, Centre on Language, Logic and Speech
GR	Institute for Language & Speech Processing (ILSP)
GR	N.C.S.R. "Demokritos"
GR	University of Patras
HU	Lóránd Eötvös University
HU	Technical University of Budapest
IT	CONSORZIO PISA RICERCHE
IT	Consiglio Nazionale delle Ricerche
IT	Fondazione Ugo Bordoni
IT	IRST
IT	Università degli Studi di Pisa
IE	Trinity College, University of Dublin
IE	University College Dublin
LT	Inst. of Mathematics & Informatics
NO	Norwegian University of Science and Technology
NO	University of Bergen
NL	Eindhoven University of Technology
NL	Foundation for Speech Technology
NL	Leyden University
NL	TNO Human Foctors Research Institute
NL	Tilburg University
NL	University of Amsterdam
NL	University of Amsterdam
NL	University of Groningen
NL	University of Nijmegen
NL	University of Twente
NL	Utrecht University
PT	Faculdade de Ciencias da Univ. de Lisboa
PT	INESC
PT	New University of Lisbon
PL	Polish Academy of Sciences
RO	Research Institute for Informatics (ICI)
SE	KTH (Royal Institute of Technology)
SE	Linköping University
RU	Russian Academy of Sciences, Moscow
UA	IRTC UNESCO/IIP
UK	Leeds University
UK	SOAS
UK	UMIST
UK	University College London
UK	University of Brighton
UK	University of Cambridge
UK	University of Cambridge
UK	University of Dundee
UK	University of Edinburgh
UK	University of Essex
UK	University of Sheffield
UK	University of Sunderland
UK	University of Sussex
UK	University of Ulster
UK	University of York

Industrial Sites

BE	Lernout & Hauspie Speech Products
BE	Lernout & Hauspie Speech Products
DE	AG für Mensch-Maschine Kommunikation mbH
DE	DaimlerChrysler AG
DE	Electronic Publishing Partners GmbH
DE	Grundig Professional Electronics GmbH
DE	IBM
DE	Langenscheidt KG
DE	Novotech GmbH
DE	Philips Research Laboratories
DE	Siemens AG
DE	Verlag Moritz Diesterweg GmbH
DE	pc-plus COMPUTING
DK	Tele Danmark
ES	Telefonica I & D
FR	Aerospatiale
FR	LINGA s.a.r.l.
FR	LexiQuest
FR	Memodata
FR	Systran SA
FR	TGID
FR	VECSYS
FR	Xerox Research Centre Europe
FI	Kielikone Oy
FI	Nokia Research Center
GR	KNOWLEGDE S.A.
HU	MorphoLogic Ltd.
IT	CSELT
IT	OLIVETTI RICERCA SCpA
IT	SOGEI
IT	Tecnopolis CSATA Novus Ortus
NL	Cap Gemini Nederland B.V.
NL	Compuleer
NL	Comsys Europe BV
NL	KPN Research
NL	Polydoc N.V.
SE	Telia Promotor
RU	ANALIT Ltd
RU	Russicon Company
UK	20/20 Speech LTD
UK	ALPNET UK Limited
UK	BICC Plc
UK	BT Research Laboratories
UK	Cambridge Algorithmica Limited
UK	Canon Research Centre Europe Ltd
UK	Ensigma
UK	Hewlett-Packard Laboratories
UK	Logica Cambridge Ltd.
UK	SRI International
UK	Sharp Laboratories of Europe Ltd
UK	Vocalis, Ltd.

What is ELSNET?

ELSNET, the European Network of Excellence in Human Language Technologies, is funded by the European Commission's Human Language Technologies programme. Members are academic and public research institutes (85) and industrial companies (50) from all over Europe.

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

Building multilingual language and speech systems requires a massive joint effort by two pairs of communities: on the one hand, the natural language and speech communities, and on the other, academia and industry. Both pairs of communities are traditionally separated by wide gaps. It is ELSNET's objective to provide a platform which bridges both gaps, and to ensure that all parties are provided with optimal conditions for fruitful collaboration.

To achieve this, ELSNET has established an infrastructure for sharing knowledge, resources, problems, and solutions by offering (information) services and facilities, and by organising events which serve academia and industry in the language and speech communities.

Electronic Mailing List

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

FOR INFORMATION

ELSNET
Utrecht Institute of Linguistics OTS, Utrecht University,
Trans 10, 3512 JK, Utrecht, The Netherlands
Tel: + 31 30 253 6039
Fax:+ 31 30 253 6000
Email: elsnet@let.uu.nl
Web: http://www.elsnet.org