

elsnews

13.1

The Newsletter of the European Network in Human Language Technologies

Spring 2004

ELSNET in FP6

Steven Krauwer, ELSNET

This is the last issue of *ELSNews* produced by our editorial team at Sussex: Lynne Cahill (who did all the editorial work) and Geoffrey Sampson. Ever since the Sussex team took over the responsibility for *ELSNews* from our main office in Utrecht with the spring issue in 2000 (under the editorship of Jenny Norris) they have managed to make it a very attractive magazine, appreciated by many members of our human language technologies community both in Europe and in the rest of the world. With a quarterly circulation of around 1300 copies to 686 sites in 49 countries worldwide, it has become an important source of information and instrument for dissemination for our whole community.

I would like to thank the Sussex team for all their efforts, and it is with great sadness that I have to say goodbye to them for now.

You would no doubt expect me to introduce our new editor here and now. Unfortunately our very successful team will not be replaced immediately. The funding contract with the EC that supported the publication of *ELSNews* has now run out, and at this moment we have not been able to secure sufficient follow-up funding to attract a new editor. This means that there will be an interruption in the publication, the first one since 1992. We anticipate that there will be no summer issue this year. After this summer we hope to be in a better position to plan and support the future of *ELSNews*. We are very sorry for this, and we can only promise you that we will do our best to overcome

our present funding dip.

As for the rest of ELSNET's operations, I am happy to announce that we have secured funding until summer 2005. We have less than we used to have (otherwise we would have been able to avoid interruptions in the publication of *ELSNews*), but we can still carry on. Uncertainties and dips in our funding situation are by no means a new phenomenon: throughout ELSNET's existence we have always been funded by the EC on a project basis, with a typical duration of between one and a half and three years.



The architectural beauty of Utrecht, home to ELSNET for the past thirteen years

Yet there is reason for some general concern: the transition from the EC's Fifth to the Sixth Framework Programme has marked a radical change in the funding instruments used by the EC, and, more importantly, also in the content of the programme. Where language and speech technology had a solid position in earlier framework programmes, they have now disappeared completely from the R&D agenda as topics in their own right. This does not imply that language and speech technology will not be addressed at all under the new programme: many of the newly started FP6 projects contain significant language or speech oriented components, and on page 13 Lynne Cahill gives us an overview of the most exciting examples. This does not take away the fact that the present programme does not favour projects that are dedicated to language

Inside this issue:

The strategic role of LRs

Nicoletta Calzolari 3

SIGDIAL

Ronnie W. Smith 5

Predicting the future: technology roadmapping

Stephan Busemann and Hans Uszkoreit 6

Opinion

Annie Zaenen 10

TALN04

Latifa Al Sulaiti 11

Speech and language in FP6

Lynne Cahill 13

The NWCL Research training programme

Bayan Abu Shawar 14

Calendar

15

Spring
2004

elsnet
.....

and speech issues alone. Fragmentation of Europe's language and speech technology efforts will therefore be a necessary consequence.

Another concern derives from the fact that the present EC policies seem to show a clear shift from 'spreading excellence' to 'boosting excellence', aimed at improving Europe's industrial competitiveness. This is in full compliance with the general objectives of the EU policies but it leads (not surprisingly) to a quite uneven distribution of EU funded R&D efforts over countries and languages. The following table, based on (provisional) participation figures in the projects emerging from the first FP6 call for proposals for the field of interfaces may serve to illustrate this. The first column gives the total number of participants from each country, the second columns lists the countries, and the third column gives the cumulative percentage.

45	France	20.9
28	Germany	34.0
26	United Kingdom	46.0
18	Italy	54.4
15	Spain	61.4
14	Switzerland	67.9
11	Netherlands	73.0
8	Sweden	76.7
7	Greece	80.0
6	Israel	82.8
6	Belgium	85.6
4	United States	87.4
4	Finland	89.3
4	Denmark	91.2
3	Ireland	92.6
3	Austria	94.0
2	Turkey	94.9
2	Slovenia	95.8
2	Portugal	96.7
2	Czech Republic	97.7
1	Senegal	98.1
1	Russian Fed.	98.6
1	Romania	99.1
1	Canada	99.5
1	Australia	100.0
215		

As said, the figures are provisional and may still change slightly, but the general pattern seems clear: a relatively small group of eight countries (one non-EU) accounts for over 75% of all participants. Please note that from this overview one should not try to draw any financial conclusions. From a language point of view it seems clear that some languages will get a lot more attention than others, justified by commercial interest rather than by inherent complexity of the language.

This is neither the time nor the place to start a discussion about whether this situation is good or bad: things are the way they are, and the only thing we, as

ELSNET, can do and should do is ask ourselves whether we can do anything to counter the problems I signalled above: the fragmentation of R&D efforts in language and speech, and the fact that some languages will be a lot better off than others in terms of coverage.

My own conclusion is that ELSNET should try to keep the language and speech R&D community together, aiming at facilitating exchange of knowledge between its members, and at porting information and expertise between languages, so that languages that receive less support from the EU programmes than others can benefit as much as possible from the huge investments the EU is making especially in the major languages.

We will use the coming months to elaborate a more detailed work programme for this, for which we hope to be able to secure sufficient funding. Close collaboration with other organisations will be one of our instruments to make our actions more effective and efficient, both during and after our present funding dip.

Our main action points at this moment are:

- improvement of our information dissemination infrastructure;
- the further construction of our roadmap for language and speech technology (more in this issue);
- our training actions, including our annual summer schools.

In connection with this last point I am very sorry to have to confirm that there will be no summer school in July 2004. Although we had (and still have) the funds for this school we were confronted with logistical problems which we were unable to solve in time. We are now investigating the practical feasibility of a (possibly shorter) school later this autumn. The 2005 summer school will take place, as usual, in July 2005. Members willing to host either of the two events should feel invited to contact us.

FOR INFORMATION

All ELSNET activities are currently coordinated by **Steven Krauwer**

Email: elsnet@elsnet.org

Web: www.elsnet.org

Offers to host summer schools, or suggestions for other coordination activities should be sent to Steven.

More information about the EU's Sixth Framework Programme can be found at fp6.cordis.lu/fp6

ELSNews is published at the Department of Informatics, University of Sussex. It is printed by the University of Sussex Print Unit.

Editor: Lynne Cahill.

Editorial Team: Geoffrey Sampson, Steven Krauwer and Amy Wall.

ISSN 1350-990X

© ELSNET 2004

FOR INFORMATION
Contributions to
ELSNews, and corrections,
should be sent to:
Steven.Krauwer@let.uu.nl



The strategic role of LRs: ENABLER and the Committee for Written LRs and Evaluation

Nicoletta Calzolari, ILC-CNR, Pisa

Language Resources (LRs) – written, spoken, and recently multimodal – are a central and strategic component of the so-called ‘linguistic infrastructure’ (the other key element being evaluation), necessary for the development of any Human Language Technology (HLT) application or product. The availability of adequate LRs for as many languages as possible is a prerequisite for the development of a truly multilingual Information Society. They play a critical role, as foundation technology, in different areas of the Sixth Framework Programme, and have been recognised as a priority within a number of national projects around Europe. The availability of LRs is also a ‘sensitive’ issue, touching directly the sphere of linguistic and cultural identity, but also having economic, cultural, social, and political implications. This is going to be even more true in the new Europe with 25 languages on a par.

The ENABLER Thematic Network of HLT National Projects in European countries – an EC funded IST project designed and started by Antonio Zampolli, with a clear strategic vision for the field of LRs – is the first broad European initiative with the mission of explicitly considering together the technical, organisational, strategic, and political issues of LRs. In ENABLER these various aspects are put together in a coherent framework, to establish medium- and long-term sets of priorities (both technical and strategic) and to promote these at national and international levels. Moreover, ENABLER has recognised the importance of promoting actions aimed at integrating the different resource types, up to now developed independently, and – as a consequence – at promoting cooperation between the communities of speech, text, and multimodality.

International research infrastructures for LRs

An important goal of ENABLER was to provide recommendations for strategic initiatives to be promoted in the field of LR production and management. Two main lines have been highlighted:

- *infrastructure initiatives* – ENABLER has promoted the creation of a new international infrastructure for linguistic resources;
- *coordination initiatives* – these concern both the national dimension and the transnational and transcontinental ones.

These lines of action seek to address the main priorities

for LRs and to define a strategy for LRs in the coming years. We highlight here some of these initiatives (for more details see the ENABLER web site).

An open and distributed framework for LRs

The need for ever growing LRs – testified also by the current US funding strategies – led us to propose and promote a change in the overall model of how to build, maintain, and share LRs. In particular, a new paradigm is required and proposed to make the web usable, i.e., an *open, distributed, and collaborative language infrastructure*, based on open content interoperability standards. Existing experience in LR development proves that such a challenge can be tackled only by pursuing – on the organisational side – a truly interdisciplinary and cooperative approach, and by establishing – on the technical side – a highly advanced environment for the representation and acquisition of linguistic information, open to the reuse and interchange of linguistic data. We are promoting the launch of a large initiative, comprising the major LR and HLT groups in Europe and world-wide, for the creation of an open and distributed infrastructure for LRs. The outcome of such an initiative could be the design of a completely ‘new generation’ of LRs.

A common Roadmap for spoken and written LRs and HLT

The workshop “International Roadmap for Language Resources” organised by the ENABLER Network in collaboration with ELSNET in Paris in August 2003 laid the foundations for the construction of a roadmap for LRs. A first list of main priorities that act as critical issues for the future of LRs was drawn up:

- provide basic LRs for a larger set of languages;
- increase multilingual LRs;
- reduce development time of LRs;
- enhance LR content interoperability;
- foster synergies between spoken and written areas and with neighbouring areas (e.g., terminology, Semantic Web);
- develop new methodologies and tools for LR man-



Nicoletta Calzolari

**Spring
2004**

elsnet
• • • • •

- agement, quick domain and application adaptation, data-driven tuning, etc.;
- an important *Declaration on Open Access to LRs* was endorsed by all participants of the ENABLER/ELSNET Workshop.

Another roadmap building meeting, as a common enterprise of the speech and language communities is to be held in conjunction with LREC 2004 in Lisbon.

Contributing to the design of overall coordination and strategy in the field of LR

International cooperation will certainly be the most important factor for the field of LRs – and consequently of HLT – in the coming years. A report produced by ELDA presents an analysis of several organisational frameworks, focussing on funding and organisational procedures for providing LRs. The pre-requisites to be addressed for the production of interoperable LRs in a cooperative framework belong to different layers: technical (specifications), validation (quality assessment), legal, commercial. In order to fill the gaps in terms of LRs cooperation on all issues combined – organisational, funding, technical, and commercial – appears to be necessary. To strengthen such cooperation, there is no doubt that an effort in coordinating this cooperation is required. A coordinated operation has already been launched in the framework of speech LRs with the creation of COCOSDA (*International Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques*). Major strategic outcomes of ENABLER with respect to international cooperation and to the design of an overall coordination strategy in the field of LRs are ICCWLRE and LangNet.

ICCWLRE

A new committee, originally conceived by Antonio Zampolli, has been established in the field of written LRs, the *International Coordination Committee for Written LRs and Evaluation* (ICCWLRE). It provides the optimal environment to continue (part of) the ENABLER mission, while, at the same time, enlarging its scope beyond European boundaries. Tasks for this committee include: information dissemination on LRs; dissemination of standards; promotion, coordination, and enabling activities; copyright and IPR; training and methodology for creation and validation of LRs; roadmaps for LRs; and political and strategic initiatives. The first joint meeting of COCOSDA and ICCWLRE is organised as a satellite event at LREC 2004, with the goal of building a roadmap for LRs, as a joint effort of the communities of speech and text, fostering future synergies among them.

LangNet

Last but not least, an initiative – LangNet – is being proposed in the framework of the ERA-Net scheme of

the Sixth Framework Programme of the European Commission to coordinate national initiatives in HLT all over Europe. LangNet is intended to provide the most natural environment to continue the efforts and the momentum gained by the ENABLER Network. Language Technologies seem to be especially well suited for the ERA-Net scheme, based on the assumption that each country wishes to conduct research activities allowing for the development of systems and applications for their language(s). It therefore seems natural that the individual countries take into account all the '(spoken and written) language-dependent' aspects and that the European Commission takes into account all the generic, 'language-independent' aspects in accordance with the principle of subsidiarity. As pointed out clearly in the final Euromap report, coordination initiatives must be put in place so as to avoid a two-speed situation, between languages which are interesting commercially, and today also politically, and those which are not (unfortunately the vast majority). This implies for Europe that coordination should be established between the European Commission and the member states and strategies should be drawn up in order to ensure a proper balance of language coverage in Europe.

The idea behind all these initiatives is to establish some sort of *permanent coordination* to capitalise on parallel existing (national or international) initiatives in the long run. The linguistic infrastructure supported by ENABLER intends to contribute to the structuring and integration of the European Research Area, addressing problems such as the fragmentation of its research base and the weakness in converting R&D results into useful economic or society benefits. To this aim, we claim it is necessary to work together and to build on many different, but related, (spoken and written) initiatives. A strategic and visionary policy for cooperation between different groups has to be debated, designed, and adopted for the next few years, if we hope to be successful. However, inside this, a realistic and stepwise approach to solving well-defined and limited aspects must be adopted. To this end, the contribution of the main actors in the field is of extreme importance. Some of the events of recent years are hopefully moving us in this direction.

FOR INFORMATION

Nicoletta Calzolari is Director of the Istituto di Linguistica Computazionale (ILC-CNR) Pisa

Email: glottolo@ilc.cnr.it

Web: www.ilc.cnr.it

ENABLER: www.enabler-network.org

COCOSDA: www.cocosda.org

LREC2004: www.lrec-conforg/lrec2004/

Initiative and Its Impact on Dialogue Management

Ronnie W. Smith, East Carolina University

One of the most important decisions that designers of a dialogue system must make concerns the level of initiative that each participant is allowed to take during a dialogue. A simple operational definition for initiative is that the conversational participant whose goals currently have priority has the initiative in the dialogue.

From a dialogue design standpoint, the simplest systems to construct are ones where the system has the initiative throughout the interaction. Such systems limit the complexity of the system by restricting the scope of user responses to the system goal, and simplify error-handling to permit the system to simply ignore out-of-context responses. Unfortunately, such systems are vulnerable to misinterpreted user responses that are in context such as interpreting the user statement, "I want to go *from* New York" as "I want to *go to* New York." This behaviour is fundamentally too inflexible for speech-based interfaces

Consequently, the next step is to design a system capable of *mixed-initiative* dialogue – dialogue where either participant may have the initiative at different points during the dialogue. The simplest extension in that regard is to allow user initiative to correct system misinterpretation of previously communicated information – even if the current system goal is to acquire other needed information.

A complicating factor in designing mixed-initiative dialogue systems is that initiative can relate to either *task goals* or *discourse goals* [1]. Task goals relate to the plan of action or plan execution that is the primary domain purpose of the dialogue, while discourse goals relate to the current conversational focus for establishing mutual beliefs (such as establishing that there exists a bad assumption in an agent's plan without actually taking task initiative to specify the next plan step).

The simple form of a mixed-initiative dialogue system, where the system retains task initiative while the user may occasionally take the discourse initiative to establish (or re-establish) proper mutual beliefs, is useful in many domains, but it is likely to be inadequate in domains where users have a great deal of expertise and/or are frequent users of a system. In extreme cases, task initiative will primarily reside with the user – for example, email processing systems. Such systems are likely to require either a very limited task model or else a sophisticated plan recognition component in order to discern the user goals.

However, the most complex system is likely to be one where users are gradually gaining expertise and task initia-

tive is more shared over time. These systems will probably require sophisticated mechanisms for user modelling, plan recognition, and generation techniques that can provide proper transition when the computer wishes to retake initiative from the user. If a user has been interested in goal Y, and the computer now wishes to take the initiative to discuss goal X, the computer cannot simply produce an utterance about goal X without relating it to goal Y. These issues largely remain open problems.



The recently published book, *Current and New Directions in Discourse and Dialogue*, provides one recent corpus-based study relating initiative to discourse structure[2]. While most of the papers in this book do not directly address issues related to initiative, in many cases the work reported could eventually be applied to that issue (e.g., annotation tools and techniques and discourse interpretation and generation techniques) while models of dialogue system development frequently assume an initiative model though it may not be a focus of study. To summarise, initiative is a pervasive aspect of discourse and dialogue study, and there is a great deal of knowledge about it that awaits discovery – particularly in the realms of multi-party dialogue (i.e., more than two participants) and non-task-oriented conversational dialogue.

FOR INFORMATION

Ronnie Smith is Associate Professor in the Department of Computer Science at the University of East Carolina

Email: rws@cs.ecu.edu
Web: www.cse.ecu.edu/~rws/

References

- [1] Jennifer Chu-Carroll and Michael K. Brown, Tracking initiative in collaborative dialogue interactions, *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, 1997, pp. 262-270.
- [2] Susan E. Strayer, Peter Heeman, and Fan Yang, Reconciling control and discourse structure, in *Current and New Directions in Discourse and Dialogue*, Jan van Kuppevelt and Ronnie W. Smith, editors, Kluwer Academic Publishers, 2003, pages 305-323.

**Spring
2004**

elsnet
.....

Predicting the Future: Technology Roadmapping

Stephan Busemann and Hans Uszkoreit, DFKI Germany

This article suggests a methodology for technology roadmapping that puts forecasting of technological developments on more systematic grounds than previous approaches. A web-based system has been developed that allows technology roadmaps to be encoded, presented to the public, and discussed widely. Within this system three roadmap documents have been encoded that highlight different aspects of the field of speech and language technologies.

Technology Roadmapping

The concept of a roadmap is a powerful and intuitive metaphor. A roadmap is a document that:

- indicates directions for a planned journey;
- shows how and in what order goals can be reached;
- indicates distances;
- is condensed into one structured presentation;
- is perspicuously presented.

A technology roadmap combines prediction of enabling developments, feasibility judgements, and scientific or economic goals into strategic planning. It puts these ingredients on a timeline, serving as a planning tool. This is feasible for periods up to five or six years. Longer periods tend to involve less reliable predictions.

Existing approaches to technology roadmapping are highly diverse. They can roughly be grouped into three categories:

Product availability plans at company websites

Using different visual means, the future developments of a product or product line are defined in terms of functionality[1].

Overview presentations, e.g., at conferences

Inquiries are evaluated that reflect experts' opinions on future applications, standards, and associated changes in society[2].

Dedicated workshops Discussions and suggestions on technology forecasting in a small group[3].

None of these approaches has adopted the roadmap metaphor for presenting information in one concise document – as a roadmap.

In 2002 the ELSNET network of excellence started building up roadmaps for the field of Human Language Technologies (HLT). We first developed a methodology of technology roadmapping that:

- identifies the types of information needed;
- systematically specifies the elements of a technology roadmap; and
- can be implemented and made available on the Internet.

Based on this methodology we developed the HLT Roadmap System and encoded several roadmaps representing workshop results that look at different aspects of HLT.

Requirements

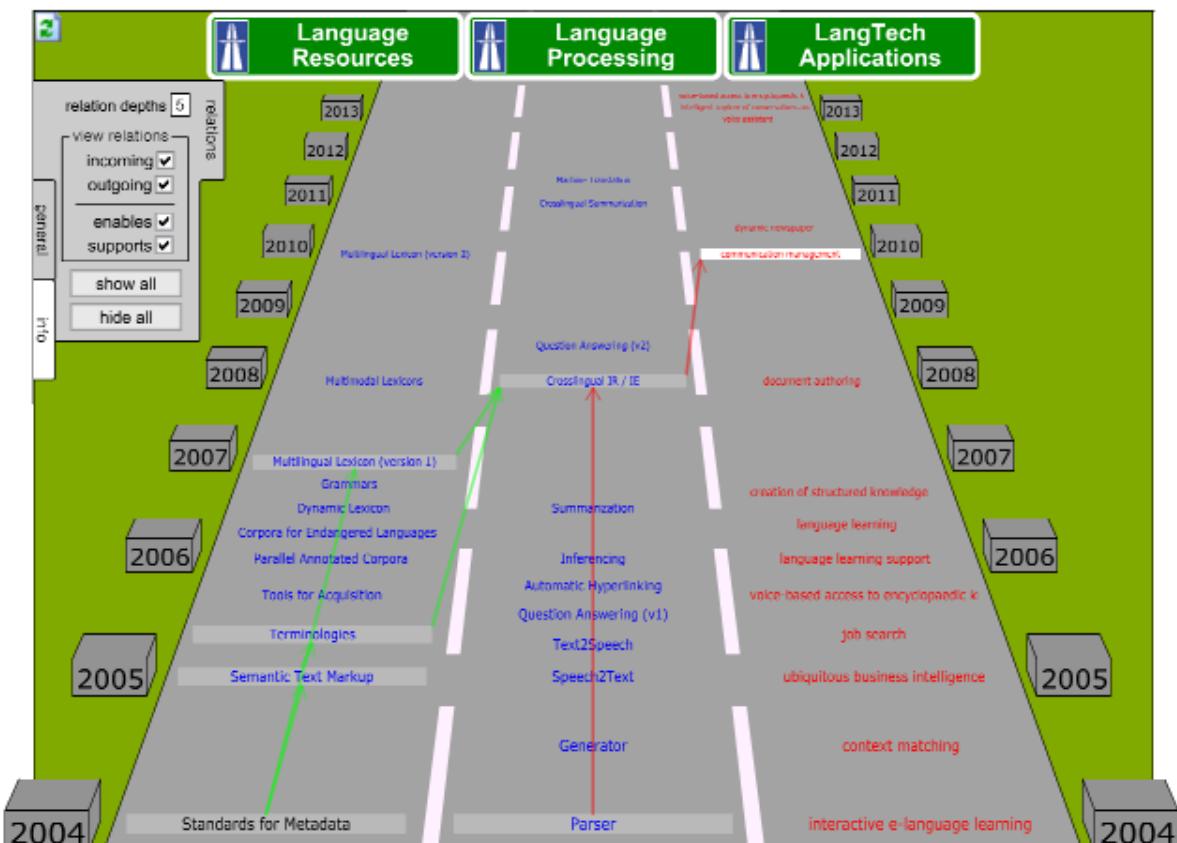
While previous approaches to technology roadmapping provide useful insights and give rise to interesting and sometimes controversial discussions, they do not aim at systematically representing technological achievements and their mutual dependencies.

We call an achievement predicted to be available at a certain point in time a milestone. Usually milestones are not precisely defined in terms of functionality and status of availability when it comes to complex topics – while it is straightforward to formulate predictions on the throughput of electronic circuits or on the growth of a morphological stem lexicon, it is difficult to predict when, for example, machine translation will be “generally useful”.

Obviously more precise milestones than that must be defined, yet different levels of granularity are needed for different prediction tasks. The usability of machine translation may well depend on progress with electronic circuits and morphological lexicons, but roadmaps may legitimately concentrate on more MT-specific milestones such as the availability of thesauri or translation memories.

The level of granularity of a presentation is, of course, limited by the need to present information in a concise roadmap document. Still, relationships with





The HLT Roadmap

the more fine-grained – and hence more basic – milestones are desirable and should be provided. This can be achieved most easily using web-based presentations; different roadmap documents can be designed in such a way that they complement each other and can be interlinked to form a comprehensive presentation.

In addition, web presentations can easily meet the need of constantly and publicly maintaining roadmap documents over time.

Usage and design principles

To design a technology roadmap document, its purpose and scope need to be determined. The HLT community and industrial consumers of that technology need to reach a common understanding of the upcoming technological development in the field and of resulting applications. A common understanding is achieved by an interactive process of contributing and discussing on the web.

The realisation of information in a technology roadmap is ontologically defined on a structural, a representational, and a presentational level.

On the structural level, we distinguish (sub)areas, milestones (subdivided into technologies, resources,

tools, applications), and relations between milestones, such as *enables* or *supports*. Milestones consist of a short name used for presentation in the roadmap, a definition (a piece of prose – see below) and an index, locating it on the timeline of a ten years period. Milestones may be related to external resources, such as other technology descriptions. Verbal comments from the public are associated with the respective piece of roadmap information.

On the representational level, the information types defined for the roadmap are represented as typed objects in a relational database. Hyperlinks to outside sources are represented as marked textual objects in the database.

At the presentational level the roadmap metaphor has rarely been taken seriously. DFKI-internal roadmaps that have regularly been submitted to the board of shareholders for about ten years form a notable exception. The presentation has been refined and taken over into the HLT Roadmap System. It consists of a three-dimensional graphical representation of a straight road with three lanes leading slightly uphill, with distance marks depicting time periods. The road is shown from a driver's point of view, from a slightly elevated angle. The major milestones are placed on the road surface.

**Spring
2004**

elsnet

They fall into different categories that are marked by colours. The lanes correspond to the main areas of technology. Lane directions are shown on a highway sign spanning the entire road.

In addition, a tabular presentation is accessible (e.g., for printing) that maintains the time axis by listing the milestones in their temporal order in three rows corresponding to the three lanes.

One of the natural visual features of the roadmap metaphor is the relative difficulty with which more distant milestones, as opposed to near ones, are recognised. This corresponds to the fact that technologies in the distant future are less precisely defined and less predictable than those in the near future. But forecasting requires us to mentally travel into the future; zooming along the time axis mimics this.

Milestones and Milestone Descriptions

Is ‘Ambient Intelligence’ a milestone? In 2010, achievements attached to this buzzword will no doubt be available. Closer thought reveals that a milestone must be more clearly identifiable. ‘Ambient Intelligence’ is rather an R&D programme than a milestone in a roadmap.

‘Machine translation’ is another example. Certainly, achievements in machine translation are clearly identifiable, e.g., as systems. Still we may ask whether ‘machine translation’ will eventually be ‘achieved’ in, say, 2010. There will be emotional discussions, but hardly well grounded positive answers. Obviously, this milestone cannot be precisely defined. What is a good milestone, then?

A milestone in the HLT roadmap should be a piece of software (in a general sense) that is clearly identifiable. Thus its description should be precise. For a parser, we should specify the output, the type of grammars it can interpret, and the speed it processes, e.g., one gigabyte of text. For a resource such as a lexicon, we should specify its size, the kind of information it encodes, and the kind of grammars it can be used with. For an application, we should specify the intended users, the functionality, and possibly the price. For all three types of milestones, information about availability and sharability should be provided.

Moreover, the state of development is important. Since the HLT Roadmap will address both consumers of technology (e.g., industry) and creators of technology (e.g., researchers and developers), the technology must be ‘ripe’ in a certain sense. This certainly means more than a validation of research results, and less than an off-the-shelf product. We might adopt the notion of *validated prototype*, to denote a piece of software for which the above infor-

mation can be given and that can, in principle, be taken over by industry for use or for professional re-implementation.

The HLT Roadmap System enforces none of the above information. It is up to the encoder of milestones to specify appropriate milestone descriptions.

Implementation

The information underlying the HLT Roadmap System is represented in a MySQL database. It is made available on the Web server through PHP scripts. The roadmap presentation is implemented with Macromedia Flash. The zooming and navigation functionalities in Flash allow for viewing close-ups.

The roadmap web site at DFKI consists of

- an introductory page,
- the selection of the roadmap to be visualised,
- the task description for that roadmap,
- the roadmap itself,
- extensive online help facilities,
- the discussion forum, and
- the password-protected administration area accessible to the roadmap and web page designers.

For any milestone in the roadmap, a menu with entries for details, comments, and relations can be opened. ‘Details’ include the short name, as seen on the road surface, the full description of the milestone, which cannot be presented on the road’s surface due to space limitations, and the category. The ‘details’ view also allows for linking to external HLT resources such as the LT-World information portal.

The ‘comments’ entry takes care of interactivity. It allows the user to send a comment about this or any other milestone to the moderated roadmap forum. The user can reply to existing messages in the forum or open up a new thread.

The ‘relations’ entry presents dependencies on other milestones. Only by establishing relationships between milestones is it possible to represent more than simple temporal sequence. Currently, relationships for enablement and support are defined. A menu on the roadmap allows the user to visualise incoming or outgoing relations for a particular milestone with a specified depth. Depth 1 shows only the direct relations, whereas by choosing a higher depth, chains of relations can be visualised.

Conclusion and Outlook

The HLT Roadmap System provides a framework for implementing technology roadmaps. So far different



aspects of HLT are looked at without claiming completeness or a wide consensus. In ELSNET, the discussion of roadmaps has so far been limited to the respective workshops. It proved difficult to have experts devote their time to a more continuous discussion and refinement of roadmap contents in order to reach a consensus document.

Therefore the ELSNET network of excellence went for a different approach: experts have been invited to prepare definitions for a 'language resources and evaluation roadmap' closely related to the information structure needed. These documents will be discussed at a joint meeting[4], and the results will be made available within the HLT Roadmap System.

Acknowledgments

This work has been funded by the European Commission under contracts IST-1999-12127 and IST-2001-34571.

We thank Thomas Eigner and Christian Federmann for implementing the HLT Roadmap System. Thanks go also to Atsuko Shimada and Melanie Siegel for encoding the 'DeepThought' roadmap, as well as to Lynne Cahill, Nicoletta Calzolari, and Steven Krauwer for their work on the 'Resources' roadmap content.

We are grateful to the participants of various roadmap workshops for their valuable contributions towards content creation.

Notes

- [1] E.g., www.3m.com/us/electronics_mfg/microelectronic_packaging/mc/tech_roadmap.html
- [2] E.g., Roger Moore's presentation at the IEEE ASRU Workshop in the Virgin Islands in 2003, www.elsnet.org/dox/moore-asru.pdf
- [3] An overview of roadmap activities can be found at www.elsnet.org/roadmap.html
- [4] *Building the Language Resources and Evaluation Roadmap*: Joint COCOSDA and ICCWLRE Meeting in conjunction with LREC 2004 in Lisbon, May 2004.

FOR INFORMATION

Stephan Busemann is the associate head of DFKI's Language Technology Lab, where he is Principal Researcher and project leader.

Hans Uszkoreit is Professor of Computational Linguistics at the University of the Saarland at Saarbrücken, Scientific Director of the German Research Center for Artificial Intelligence (DFKI), and Head of DFKI Language Technology Lab.

Email: busemann@dfki.de; uszkoreit@dfki.de
Web: www.dfki.de

The HLT Roadmap can be found at elsnet.dfki.de.

The LT-World information portal can be found at www.lt-world.org.

ing there and it has survived on various grants mainly manned by student researchers. Student researchers are typically, and justifiably, more interested in doing something new than in ironing out small problems and improving performance. (For the results obtained by one of the HPSG endeavours see [1].) The European Commission has funded some research, most prominently the Aleph project, but it hasn't had the staying power to see such a project through across funding cycles and has put progressively more and more pressure on the projects it funds to achieve short-term results.

In the current climate, then, one cannot find a funding agency willing to bet on an integrated team working for a reasonably long time and nobody in his/her right mind would dare to start a similar effort, except maybe at Microsoft. Full parsing and generation are now possible, but it might have become impossible to solve other problems that need a concentrated effort.

Note from the editor:

I would like to thank Annie for contributing five fascinating columns over the past year and a bit. It has been a pleasure to work with such an amenable and prompt contributor!

contd from p. 10

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] Baldwin, Tim, H. Beavers, E. M. Bender, D. Flickinger, A. Kim and S. Oepen. 2004. Beauty and the Beast: what running a broad-coverage precision grammar over the BNC taught us about the grammar and the corpus. In *Proceedings of the International Conference on Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives*. Sonderforschungsbereich 441,
- [2] Riezler, S., T.H. King, R.M. Kaplan, R. Crouch, J.T. Maxwell III, and M. Johnson. 2002. Parsing the Wall Street Journal using Lexical-Functional Grammar and Discriminative Estimation Techniques. *ACL*, Philadelphia.
- [3] Maxwell III, J. T. and R. M. Kaplan. 1991. An overview of Disjunctive Constraint Satisfaction, in M. Tomita editor, *Current Issues in Parsing Technology*, Kluwer, Dordrecht. Also in Dalrymple et al., *Formal Issues in LFG*, CSLI publications.

Spring
2004

elsnet
.....

...But full parsing is impossible

Annie Zaenen, PARC

Over the last five years, I don't know how often I have read a clause regretting the impossibility of full parsing and expressing the conviction that one has to make do with less desirable approximations. When this mantra was first invented it might have expressed a truth, but by now the reality is different and efficient full syntactic parsing is quite possible.

In what follows I will answer some of the frequently cited problems with full parsing by giving information about the system I know the best – XLE, developed at PARC [2] – but I do not want to imply that it is the only one that can deliver satisfactory performance.

Full parsers are slow. Compared to POS taggers this is true, but they are not too slow. The XLE parser in its ‘CORE’ version (excluding parses that are dispreferred in the full version) takes about 0.54 seconds to parse a sentence from the PARC dependency bank with an F-score of 77.6. (The full system gets an F-score of 79.6 but takes 1.76 seconds per sentence). This might not be quick enough if one wants to consume the whole web but it is good enough for a lot of applications.

Full parsers yield so many parses that no application can cope with them. No, the hand-written XLE grammar can be run with a log-linear disambiguation component. In this mode, it outputs only the most probable parse. It can, however, also be run without this component when one does not want to eliminate parses too early. Very general ambiguity management techniques (see [3] for the basic idea) allow one to put off the resolution of ambiguities until the appropriate moment. The best moment

depends on the application. If the system has only syntactic information, ambiguities are best resolved when the syntactic parse is done, but if there is



Annie Zaenen

a knowledge representation component with real world information it might be best to put it off until that information can be taken into account.

Full parsers are brittle. No, the XLE parser allows fragment-parses so all sentences yield a result. On the *PARC 700 dependency corpus*, with stochastic disambiguation, we get an F-score of 82.5 for the 74.7% of the sentences that were fully parsed and an F-score of 69.0 for the 25.3% that were only partially parsed. (Note that these results are for the stochastically chosen parse, whether or not a correct parse is among the many possible.)

The system can be configured in various ways. It can take the output of a finite state pre-parser to speed up the parsing further. Ambiguities can be resolved at different points. Only the most likely parse can be rendered or all parses can be made available. Optimality marks can be used to calculate the preferred parse or not. So, in a lot of cases, you can have your cake and eat it too.

Moreover, the general architecture is reversible: it can not only parse but also generate and is linked to a general rewrite system, that can be used for translation and other applications, and a semantic analyser, that can interface to a knowledge representation system.

Together with parsing efficiency, grammar writing has also been streamlined and the PARC researchers estimate they can develop a grammar for a new language in two person years. Development can be quicker when new grammars bootstrap off already developed ones (a technique also used by the Microsoft research team).

These results did not fall out of the sky. This unabashed boasting is only possible because the system was developed steadily with a small but stable team over twenty years. This was possible when the political climate valued long-term fundamental research. Xerox kept the effort going over this whole period, although not always without grumbling. Other efforts that were started around the same period have not had such clear results, maybe in part because they didn't find the right team, but surely more because they have not had the same type of support. An HPSG effort, for instance, was started at Hewlett Packard but soon lost its fund-

contd on p. 9

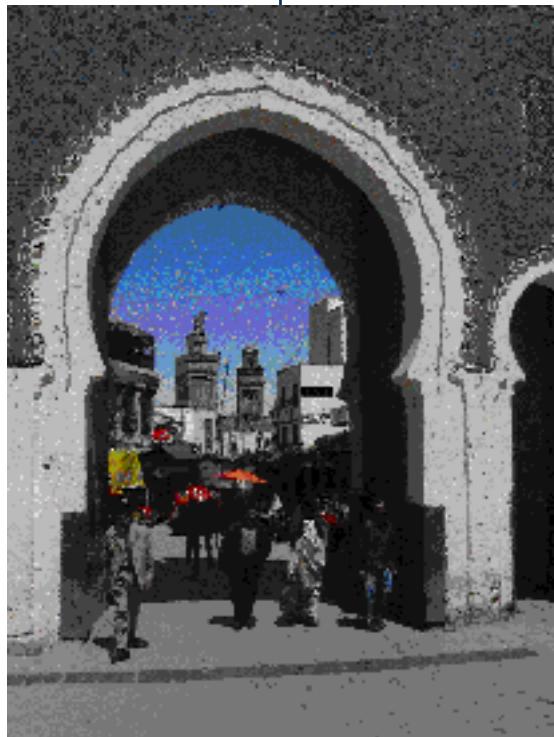
TALN'04: The North African Experience

Latifa Al Sulaiti, University of Leeds

TALN'04 (*Traitement Automatique du Langage Naturel*) in conjunction with JEP 2004 (*Journées d'Etude sur la Parole*) and RECITAL 2004, the conference for young researchers – is an annual conference, which has been held since 1994. This year it was held for the first time outside France. It was organised by the Université Sidi Mohamed Ben Abdellah (Fes-Maroc), Ecole Normale Supérieure (Fes-Maroc) and Laboratoire Parole et Langage (CNRS-Université de Provence). The conference included oral and poster presentations, invited speakers, workshops, and tutorials. The conference took place between 19th and 22nd April in Fez at the ‘Palais des Congrès’ in the new town which is not far from the main attraction: the labyrinthine old town with its spectacular features.

There were around 270 participants of which 77% were from France and the rest were from North Africa, North America, and other European countries. Most of the papers were therefore delivered in French. The papers for TALN'04 covered topics on semantics, discourse, parsing, summarisation, dialogue information retrieval, and machine translation. New projects and software presented at this conference included development of an electronic dictionary for Greek; a multilingual summarisation system which generates summaries in three languages; a system for evaluation of automatic summarisation; an automatic end-of-book indexing system; a new approach to information retrieval, which incorporates pairs of words without constraint on word order; an automatic anonymiser to be used on the web for legal document; and a system for corpus processing. The languages handled in these projects are French, German, Chinese, Korean, Italian, Spanish, Greek, English, and others. JEP04 also demonstrated a wide range of topics on speech in several languages including some Arabic dialects Berber, and several other European languages. Some of the projects presented

and worth mentioning are: EVALDA, which is developed to establish an evaluation infrastructure for text and speech processing systems, and ESTER, which has the aim of evaluating broadcast news transcription systems. Both have been developed for French[1]. There were numerous papers that are worth reporting, but it is impossible to go through all of them in this short report. For those who are interested, the proceedings are available online at the conference web site (below).



One of the main gates in the old town or Medina

The conference had a special session for Arabic research: ‘Arabic language processing, text and speech’, which was relevant to my own research interests. It contained several sessions: plenary speech sessions, parallel speech sessions, and two poster sessions. The sessions covered a number of topics, among them text-to-speech synthesis, morphology, text summarisation, information retrieval, machine translation, Arabic corpora and resources, and others. Some presenters of posters gave demonstrations of programs developed for morphological analysis, syntactic analysis, and automatic diacritisation (presented as text mode and speech mode). It was pleasing to see that research on Arabic NLP is dealing with wider areas of applications and gradually making significant progress. This session began with the invited speaker Khalid Choukri, the executive director of ELRA. His talk, which was titled “Présentation d’activités linguistiques” focused on four main points: the ELDA agency and its resources Technolangue projects, the nature of Arabic language processing, and the announcement of the Nemlar conference in Egypt and its significance. Papers presented in the Arabic session are accessible at the URL [2]. The final day of the conference was reserved for workshops and tutorials which brought participants closer in terms of exchanging ideas and cooperation.

Conference report

Spring
2004

elsnet

In addition to the activities in the lecture rooms, participants enjoyed occasional visits to the old town. They further enjoyed great hospitality including coffee breaks with the best of Moroccan gateaux, refreshing mint tea, and lunch offered every day in the nearby four-star hotel Volubilis. On the third day of the conference the participants enjoyed a big dinner party in 'Al Firdaous' restaurant in the old town (medina). In my view it was a truly fascinating experience, which offered the combination of an impressive ornate palace exhibiting the greatness of Arab architecture, numerous delicious typical Moroccan dishes, and a variety of Moroccan folk music such as Gnaoua and Andalusian music.

In general, I think the organisation of the conference was excellent. For this Malek Boualem, Philippe Blache, Noureddine Chenfour, Bernard Bel, and the other local organisers deserve a big "Thank You!"



The souk (market) in Fez

contd from p. 13

content. What we must now hope for, as stressed by Steven Krauwer on the front page, is that the various efforts across the EU (and outside – many of the participating institutions are non-EU) do not lead to a fragmentation of the speech and language community.

A glance through the synopses of these projects suggests significant scope for combined effort, and the networks of excellence should ensure that cross-discipline and cross-national research continues to thrive.

FOR INFORMATION

For more information about the EU Sixth Framework Programme and details of the different types of project, see: [/fp6.cor-dis.lu/fp6](http://fp6.cor-dis.lu/fp6)



Participants at the poster session

FOR INFORMATION

Latifa Al Sulaiti is a PhD student in the School of Computing at the University of Leeds

Email: latifa@compleeds.ac.uk

Web: www.compl.leeds.ac.uk/people/rs/latifa.html

Conference web site (proceedings):
www.lpl.univ-aix.fr/jep-taln04/proceed/

[1] www.technolangue.net/article60.html

[2] www.lpl.univ-aix.fr/jep-taln04/proceed/actes/arabe.htm

The projects mentioned here have the following co-ordinating organisations:

AMI – University of Edinburgh, UK

CHIL – Fraunhofer Institute for Information and Data Processing, Munich, Germany

DIVINES – MUL TITEL ASBL, Mons, Belgium

ENACTIVE – Scuola Superiore di Studi Universitari e di Perfezionamento Santa Anna, Pisa, Italy

HIWIRE – Thales Avionics, Massy, France

HUMAINE – The Queen's University of Belfast, UK

MWEB – Geie Ercim W3C, Biot, France

PASCAL – University of Southampton, UK

SIMILAR – Université Catholique de Louvain, Belgium

TAI-CHI – University of Wales, Cardiff, UK

TALK – Universität des Saarlandes, Saarbrücken, Germany

TC-STAR – Istituto Trentino di Cultura, Trento, Italy

T'N D – Politecnico di Milano, Italy



Speech and language in FP6

Feature

Lynne Cahill, ELSNews

The new projects due to commence under the European Union's Sixth Framework Programme are, as already mentioned in Steven Krauwer's contribution to this issue, rather different in nature from the previous European funded projects. The commission has chosen to fund a smaller number of much larger projects, under the auspices of which groups from diverse academic and national backgrounds will work together under the direction of a single institution.

There is a perception in the speech and language community that our field has been relatively neglected in the current funding round. Whether or not this is the case, there are a number of interesting projects with greater or smaller speech and language elements. Here we present some of those projects. As these projects are still in the planning stages, some of the details are sketchy.

Probably one of the most exciting looking projects is the **AMI** integrated project, led by the University of Edinburgh. It is based on a premiss not unlike the VerbMobil project, namely "smart meeting rooms" and "remote meeting assistants". The areas that will be addressed within this project include (multilingual) speech signal processing, multimodal dialogue modelling, summarisation, and annotation of multimodal recordings.

CHIL (also an integrated project) stands for Computers in the Human Interaction Loop, and it aims to address questions of how computers can assist in human-human interaction, as opposed to the more traditional human-computer interaction.

Several of the projects cover issues of Human Computer Interaction, some with explicit speech or language elements, others not. For example, the **HUMAINE** and **ENACTIVE** networks of excellence each address slightly different aspects of extralinguistic information within HCI. HUMAINE will address questions of emotion, with the aim of providing "emotion-oriented systems". EN ACTIVE aims to produce "sensible machines", providing a generic interface between the human senses and the machine.

The **SIMILAR** network of excellence aims to merge research on HCI with research on signal processing and thus involves speech groups as well as groups working on HCI. It is one of the largest of the projects involving speech and language and will establish an international journal, run summer schools and sessions at international conferences.

The **T'N D** and **MWEB** projects only marginally involve speech and language work, addressing as they do multimodal access to systems with no explicit involvement of language. The T'N D project aims to provide an interface for designers that will enable them to use manual manipulation directly, while MWEB (a specific support action) aims to provide multimodal web access making use of the technological advances in mobile communication to provide an alternative to current internet technology.

In the field of speech, there are two very significant and, on the face of it, quite similar projects **DIVINES**, which is a network of excellence, and **HIWIRE**, a STREP (Strategic Research Project), both aim to improve robustness in speech recognition. HIWIRE has the explicit goal of improving robustness to a level that will permit the use of ASR in aeronautical applications, including voice input by pilots in the cockpit and use of PDAs and other mobile technology in aeronautical environments. DIVINES, on the other hand, has the more general aim of improving robustness of ASR by developing improved feature extraction and modelling at both the acoustic and lexical levels. The **TAI-CHI** project is of marginal interest to the speech community, addressing issues of acoustic signal processing in "tangible interfaces".

The **TC-STAR** integrated project also involves speech, addressing speech to speech translation. In addition to the main overall aim of improving speech to speech translation to "significantly reduce the gap between human and machine performance", the project will implement an evaluation infrastructure (to enhance competition) and a technological infrastructure (to enhance delivery of results).

The **PASCAL** network of excellence focuses on machine learning, with application in a number of different fields, including both speech and language. The project description explicitly refers to speech, natural language processing, information retrieval, and textual information processing.

Finally the **TALK** project addresses dialogue management, proposing to bring together the TRINDI and SIRIDUS approaches to dialogue management and develop them to cover multimodal as well as multilingual dialogues.

As we can see, there are a number of very interesting looking projects, with significant speech and language
contd on p. 12

Spring
2004

elsnet
.....

The NWCL Research Training Programme 2004

Bayan Abu Shawar, University of Leeds

The North West Centre for Linguistics was launched in November 1997 with the aim of developing a regional centre for linguistics research. Currently, it involves the universities of Lancaster, Liverpool, Manchester, Salford, and Wales Bangor as well as UMIST and Edge Hill College of Higher Education, who all have representatives on the steering committee. The NWCL organises different events such as: conferences, training sessions, and annual lectures.

The NWCL Research Training Programme for Spring 2004 was the fourth in a series of training programmes on linguistic and computational linguistic topics that have been taking place annually since 1999.

This year the event was held in University of Manchester from 29th March to the 2nd April. The programme was partitioned into three sessions: Comparative Syntax, Computational Linguistics, and afternoon sessions which included various lectures that need three hours. The lecturers came from different universities including: Manchester, Lancaster, Essex, Amsterdam, UMIST, Leeds, and Oxford.

The one week programme was full of lectures. It is difficult, therefore, to review in detail all the topics that were covered. However, in these comments I will mainly touch on lectures that related to my work and interests in using the corpus to generalise a 'chatbot' system.

On the first day Paul Bennett gave an overview of Computational Linguistics. The talk started by listing the benefits of using computational tools to help linguists in analysing, understanding, and translating large texts. Then he illustrated the difference between techniques for processing language by computer and applications of those techniques. Various areas of CL, such as morphological analysis, tagging, chunking, and parsing were discussed.

Debora Field tackled parsing techniques in more detail. She defined parsing as the process of deriving the structure of

a sentence. Three classes of parsing process were described: full parsing covering every part of the input text; partial parsing where parts of the input text are parsed fully, but other bits are completely ignored; and finally shallow parsing, which parses the whole input text 'badly'. Debora discussed two methods to do parsing: bottom-up and top-down techniques. On the other hand she distinguished between NL generation and NL analysis machines. Natural language generation systems generate natural language output for (usually) non-linguistic input. An early example of this is the ELIZA chatbot system, which simulates a psychotherapist and interacts with users turn by turn using natural language. Inside ELIZA there are no linguistic formalisms at all. All that happens is that ELIZA takes the user's input, and uses that to generate NL output. In contrast NL analysis systems take some NL input and produce a parse tree, parts of speech (POS tagging), or produce a representation of the meaning using an interlingua.

Jeanette Sakel from University of Manchester presented an interesting lecture about fieldwork. The lecture sketched an overall philosophy and methodology of fieldwork, focusing principally on the practical issues of documenting languages and writing grammars. She described her fieldwork experience of the Moseten language of Bolivia. The first step in documenting a language is making contact with speakers, and recording the text. Then it is necessary to analyse the text, eliciting the words, the structure, and the translation. The final step is to check with several consultants.

In the same context Peter Austin emphasised that the documentation of language is useful for research into the socio-cultural life of the community and must be analysed and processed in such a way that it can be understood by researchers of other disciplines, as well as the speech community.

Overall, the NWCL training week was a nice opportunity to see the work of other researchers in different universities, and to chat directly with them. The participants enjoyed the friendly, familiar environment.

FOR INFORMATION

Bayan Abu Shawar is PhD student in the School of Computing at the University of Leeds

Email: bshawar@comp.leeds.ac.uk

Web: www.comp.leeds.ac.uk/people/rs/bshawar.html



Participants on the programme

Calendar

Future Events

- Jun 6-11** *Fourth Annual Workshop in Lexicography and Lexical Computing:* Brighton, UK
 Email: inquiries@lexmasterclass.com URL: www.lexmasterclass.com
- Jun 7-11** *Categorial Grammars 2004:* Montpellier, France
 Email: degeilh@lirmm.fr URL: www.lirmm.fr/CG2004
- Jun 19-21** *CATALOG'04: Eighth Workshop on the Semantics and Pragmatics of Dialogue:* Barcelona, Spain
 Email: catalog04@upf.edu URL: www.upf.edu/ca talog04
- Jun 21-25** *NASSLLI04 (North American Summer School in Logic, Language and Information 2004):* Los Angeles, USA
 Email: nasslli04@humnet.ucla.edu URL: www.linguistics.ucla.edu/nasslli04
- Jul 6-10** *EURALEX 2004:* Lorient, France
 Email: elx2004@univ-ubs.fr URL: www.univ-ubs.fr/euralex2004
- Jul 14-16** *Third International Conference on Natural Language Generation:* New Forest, UK
 Email: inlg04@itri.brighton.ac.uk URL: www.itri.brighton.ac.uk/inlg04
- Jul 21-26** *42th Annual Meeting of the Association of Computational Linguistics:* Barcelona, Spain
 Email: acl@aclweb.org URL: www.acl2004.org
- Aug 9-20** *ESSLLI 2004: 16th European Summer School in Logic, Language and Information:* Nancy, France
 Email: esslli_helpdesk@loria.fr URL: esslli2004.loria.fr
- Aug 23-29** *COLING 2004:* Geneva, Switzerland
 Email: hess@cl.unizh.ch, URL: www.issco.unige.ch/coling2004

Submission deadlines

- Jun 1** *ICDM '04: The Fourth IEEE International Conference on Data Mining,* Brighton, UK
 November 1-4, 2004, Email: icdm@wi-lab.com URL: icdm04.cs.uni-dortmund.de
- Jun 30** *Artificial Intelligence Journal: Special Issue on “Connecting Language to the World”,*
 Email: aereiter@csd.abdn.ac.uk, URL: www.csd.abdn.ac.uk/~ereiter/langworld

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

Spring
2004

elsnet
.....

ELSNET

Office

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

Task Groups

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

Linguistic & Speech Resources

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

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

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

The ELSNET Participants: Academic Sites

A	University of Vienna	I	Università degli Studi di Pisa	D	DaimlerChrysler AG
A	Austrian Research Institute for Artificial Intelligence (OFAI)	I	Consorzio Pisa Ricerche	D	Langenscheidt KG
A	Vienna University of Technology	IRL	Fondazione Ugo Bordoni	D	Verlag Moritz Diesterweg GmbH
B	University of Antwerp - UIA	LT	IRST	D	aspect Gesellschaft für Mensch-Maschine Kommunikation mbH
B	Katholieke Universiteit Leuven	NL	Consiglio Nazionale delle Ricerche	D	Philips Research Laboratories
BG	Bulg. Acad. Sci.-Institute of Mathematics and Informatics	NL	Trinity College, University of Dublin	D	Grundig Professional Electronics GmbH
BY	Belarussian Academy of Sciences	NL	University College Dublin	D	Acodata GmbH
CH	SUPSI University of Applied Sciences	NL	Inst. of Mathematics & Informatics	D	IBM Deutschland
CH	University of Geneva	NL	Foundation for Speech Technology	D	Varits Communications
CZ	Charles University	NL	University of Twente	D	Tele Danmark
D	Universität des Saarlandes	NL	University of Groningen	DK	SchlumbergerSema sæ
D	Ruhr-Universität Bochum	NL	Tilburg University	E	Telefónica I&D
D	Universität des Saarlandes CS-AI	NL	Eindhoven University of Technology (TUE)	E	Xerox Research Centre Europe
D	German Research Center for Artificial Intelligence (DFKI)	NO	University of Nijmegen	EL	KNOWLEGDE S.A.
D	Institut für Angewandte Informationsforschung	NO	Leiden University	F	LINGA s.r.l.
D	Universität Erlangen-Nürnberg - FORWISS	P	Utrecht University	F	Systran SA
D	Universität Hamburg	P	Netherlands Organization for Applied Scientific Research TNO	F	Xerox Research Centre Europe
D	Christian-Albrechts University, Kiel	PL	University of Amsterdam (UvA)	F	Memodata
D	Universität Stuttgart-IMM	RO	Norwegian University of Science and Technology	F	Aerospatiale
DK	University of Southern Denmark	RU	Technology	F	VECSYS
DK	Center for SpråkTeknologi	S	University of Bremen	F	SCIPER
DK	Aalborg University	S	Romanian Academy	F	TGID
E	Universidad Politécnica de Valencia	UA	Russia n Academy of Sciences , Moscow	I	Kielilohne Oy
E	University of Granada	UK	IRTC UNESCO/IIP	I	No kia Research Center
E	Universidad Nacional de Educación a Distancia (UNED)	UK	University of Lisbon	FIN	MorphoLogic Ltd.
E	Polytechnic University of Catalonia	UK	New University of Lisbon	HU	OLIVETTI RICERCA SCpA
E	Universitat Autònoma de Barcelona	UK	Polish Academy of Sciences	I	LOQUENDO
EL	National Centre for Scientific Research (NCSR) 'Demokritos'	UK	Russian Academy	LV	TILDE
EL	University of Patras	UK	Linköping University	NL	Compuleer
EL	Institute for Language & Speech Processing (ILSP)	UK	IRTAC UNESCO/IIP	NL	Knowledg Concepts BV
F	LORIA	UK	University of Edinburgh	NL	Sopheon NV
F	Inst. National Polytechnique de Grenoble	UK	Leeds University	PL	Neurosoft Sp. z o.o.
F	LIMSI/CNRS	UK	University of Sheffield	RU	Russian Company
F	IRISA/ENSSAT	UK	University of Essex	RU	ANALIT Ltd
F	Université Paul Sabatier (Toulouse III)	UK	University College London	S	Sema Infodata
F	Université de Provence	UK	The Queen's University of Belfast	S	Telia Promotor AB
GE	Tbilisi State University, Centre on Language, Logic and Speech	D	University of Brighton	UK	Vocalis Ltd.
HU	Lóránd Eötvös University	D	University of York	UK	Imagination Technologies plc
HU	Technical University of Budapest	D	UMIST	UK	Hewlett-Packard Laboratories
			University of Dundee	UK	Canon Research Centre Europe Ltd
			University of Ulster	UK	ALPNET UK Limited
			University of Cambridge	UK	Reuters Ltd
			University of Sussex	UK	SRI International
			University of Sunderland	UK	Shap Laboratories of Europe Ltd
				UK	BT Adastral Park
				UK	Logica Cambridge Ltd.
				UK	20/20 Speech LTD

Industrial Sites

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

Subscriptions

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: www.elsnet.org

