

elsnews

7.2

The Newsletter of the European Network in Language and Speech

April 1998



Will there be life under FP5?

Steven Krauwer, ELSNET coordinator

Will there be life under FP5? This may look like a silly question: of course life will not stop under the Commission's Fifth Framework Programme. But if we narrow the question down to 'Will there be life *for ELSNET* under FP5?', the answer is less obvious. After all, technically speaking ELSNET is no more than one out of hundreds of EC projects, with a project number, a project officer, a contract, and an expiration date.

But I don't think that that's the right way of looking at an enterprise like ELSNET, and I am sure that when the Commission initiated the Networks of Excellence, they had something broader in mind than just starting another series of projects. In my view, ELSNET is a community, united by a common long-term goal (and hence common needs and problems), and populated by a variety of people, each with their own views, skills, interests, dreams and backgrounds. The main question concerning our life under FP5 is therefore whether our common goal will have enough relevance under the next framework programme to justify the continued existence of our community. And, if so, whether we will have to adjust our way of life to the changes that may take place around us.

One quick look at any document on FP5 produced so far makes it clear that multilingual integrated language and speech systems will still be a top priority. The need for such systems is still there, and so are the problems: while some have disappeared, at least as many new ones have come up. But there does seem to be a change of context. Where the current programme is best described as industry-driven, with its main focus on commercial applications, the next framework programme seems to be much more driven by the needs of individuals and society. It will put special emphasis on two types of communication processes: from human to human, and from human to machine or information system. Both types of communication give rise to new inspiration and new challenges. Let me mention just two typical examples, which interestingly enough both seem to point in the same direction: multilinguality and interfaces.

Although multilinguality is not the only obstacle for inter-human communication, it is a very important and very visible

one, especially in the ever expanding European Union. It is very tempting to infer from this that we should put our efforts into more, better, and bigger machine translation systems, and I am certainly not arguing that we shouldn't do this. But the observation that translation is just one of many ways of tackling the multilinguality problem seems to be much more interesting. Multilingual authoring and generation (of text and speech) and translators' aids have already been the main themes of a number of significant projects funded by the EC and national governments under the current programme. I expect that under the next framework programme we will see new approaches, where modalities (written language, speech, gestures, pictures, sounds, facial expressions, and so on) will be combined in order to overcome language and communication barriers.

Interfaces for facilitating communication between humans and machines appear to be moving in the same direction. Buttons and keyboards are already making place for graphical and speech interfaces, but I don't believe that the future will be dominated by text or speech only interfaces, or even by combinations of the two. It will bring us interfaces with machines or information systems that allow for maximal ease and naturalness of interaction, by using many different combinations of modalities, just as human beings do when they communicate,

What does this mean for ELSNET? One conclusion is obvious: the problems that justified ELSNET's creation back in 1991 are still there, and still relevant. Another conclusion might be more painful: where in the past we may have thought that bringing the language and speech communities together was the main sociological or cultural challenge for the network, it is now becoming clear that just the two of them will not be enough to meet the challenges offered (or rather: imposed) by the new framework programme. I don't think this means that ELSNET's days are over, but I do think that from now on we can no longer afford to think of ourselves as a two — now halfway integrated — communities, living nicely together in their own cocoon. We're going to be part of something bigger, and I hope and expect that ELSNET's future actions will reflect this. So, to get back to the question we started with: Will there be life under FP5? I believe that there is only one answer: yes, there will be life — and it will be more exciting than ever.

Programmes,
actions, main lines
and challenges 2
An outline of FP5

What's in it for us? 3
Giovanni Varile on
Human Language
Technologies in
FP5

Making
connections 5
Afterthoughts on
Eurospeech'97

Lessons learnt from
the Fourth
Framework 6
Interview with
Brian Oakley

Wonderland in
sight? 8
Report on the
ELSNET in Won-
derland conference

Let's Twist Again 9
Marc Blasband
proposes a TWIST

EB Minutes 10

Future Events 11

April 1998



The Fifth Framework Programme is taking shape

*The European Union organizes its spending on scientific research in five-year plans called **Framework Programmes**. We are currently reaching the tail end of the **Fourth Framework Programme**, and discussions on the **Fifth Framework Programme (FP5)** have been underway for a few years now. FPV covers science, research and development for the period 1999-2004.*

*On February 12th, the Council of Research Ministers reached an agreement on the budget: 14 billion ECU will be available for FP5. This is not as much as some had hoped for; indeed, **Edith Cresson**, Commissioner for research, education and training, described the agreement as "a dark day for European research".*

The agreement is by no means the final step in the long process of getting FP5 approved by the European Parliament and the Council. Political and financial discussions will continue, but FPV looks set to be on target for a first call for proposals early in 1999, and for a first set of projects to start in the Autumn of 1999.

While the specific content of the new Framework is still being discussed, its general outline is becoming clearer. We know, for instance, that FP5 will have four **Thematic Programmes** (see right column).

The second Thematic Programme is commonly referred to as the **Information Society Technologies (IST)** Programme. Within IST there are four **Key Actions** or focused RTD areas. These Key Actions are problem-oriented, and will cover large and small applied, generic and basic research projects.

Work related to **Language Engineering** can be situated in many of the key actions. For example, the first action, 'Systems and services for the citizen', includes computerized clinical systems, communication tools for people with special needs, and intelligent multi-functional systems to facilitate interactions between communities and administrations. But the bulk of LE work will be carried out under the third key action, '**Multimedia content and tools**'. This action has four main lines.

In July 1997, the Commission produced a discussion document on **Human Language Technologies**. Essentially, this document proposes that research and technological development in Human Language Technologies should be focused on a small number of challenges to which it can contribute in a significant way.

*On the next two pages, **Giovanni Varile** discusses these challenges, and the place of **Human Language Technologies** in the **Fifth Framework Programme**, in more detail.*

FP5 Thematic Programmes

1. Improving the quality of life and the management of living resources
2. Creating a user-friendly information society
3. Promoting competitive and sustainable growth
4. Preserving the ecosystem.

FP5 Key Actions

1. Systems and services for the citizen
2. New methods of work and electronic commerce
3. Multimedia content and tools
4. Essential technologies and infrastructures.

Multimedia content and tools: main lines

1. Interactive electronic publishing
2. Education and Training
3. Human Language Technologies
4. Advanced technologies

Main challenges for Human Language Technologies in FP5

1. Active content
2. Natural interactivity
3. Multilinguality

ELSNNews is published at the Centre for Cognitive Science, University of Edinburgh using Aldus Pagemaker™. It is printed on recycled paper by Lutton Press, Ltd.

Editors: Mimo Caenepeel, Mariken Broekhoven and Steven Krauwer.

Lay-out and production: Mimo Caenepeel

ISSN 1350-990X

© ELSNET 1998.

FOR INFORMATION Contributions to ELSNNews, and address corrections, should be sent to:

elsnews@let.ruu.nl

Tel: +44 131 650 4594

Fax: +44 131 650 6626

Material for the next issue is due:

15 May 1998

FOR INFORMATION

All documents related to the Fifth Framework Programme can be found on <http://www.cordis.lu/fifth/src/docs.htm>

The discussion document on Human Language Technologies (July 1997) can be found on <http://www2.echo.lu/langeng/en/fp5/lt.html>

Human Language Technologies in FP5

Giovanni Battista Varile, DG XIII E 5, European Commission

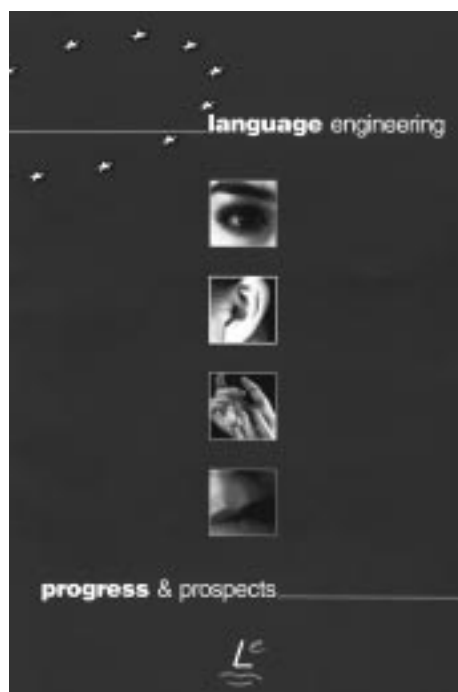
Human Language Technologies (HLT) will be one of the activities within the key action on **Multimedia Content and Tools** of the programme on **Information Society Technology (IST)** in FP5. The IST programme is the successor to the three **Information Communication Technologies (ICT)** programmes **Esprit**, **Telematics Applications** and **Acts**. Key actions are the problem-oriented activity pools around which specific programmes within the Fifth Framework Programme are organized.

As the preparation of future HLT activities is approaching its final phase, the time is right for providing an update on the progress made so far.

The Consultation Process

Consultation with a large number of key actors has been the main source of input for the definition of future HLT activities. Many *ELSN* readers will remember that consultations started in June 1996 and continued in the second half of that year. In the first half of 1997 we organized a number of focused meetings with representatives of providers and user organisations, and with the research community. Discussions with the Language Engineering working party of our programme committee have provided further input.

The result of this process was a working document which has been available for consultation and comments since July 1997 (for web-site address see the information box at the end of this article). Consultation with the community at large and Member States will continue this year and during the implementation of the programme.



A detailed report (120 pp) on *Language Engineering in FP4*, is available. To order see the information box at the end of this article.

Hereafter, I will provide an overview of the key features of future HLT activities, reflecting the consultations carried out so far.

The Approach

Language technology is a key Information Society enabler. The 'commodization' process that ICT solutions have to go through in order to realize the Information Society will become feasible through the deployment of language technologies. The language intelligence thus built into the infrastructure will empower people to fully participate in the Information Society, and enable businesses to fully exploit the new opportunities offered by the information age.

Because of this it was decided to focus HLT on global challenges central to key drivers of the Information Society, to which HLT can contribute in a substantial way. The drivers in question are the Web, high bandwidth digital communications, the resulting globalization of society and business, and the convergence of communication, computing and (multi-)media.

Global Challenges and HLT Answers

- Making **content active** in order to ensure a maximally efficient use of digital information — Web and multimedia content in particular — and assist an effective assimilation of knowledge for all.

This requires language processing models and techniques, including deep information analysis, knowledge extraction, summarization, meaning classification and structuring and metadata generation.

- Support **natural interactivity** in technology mediated interpersonal communication — collaborative services in particular — and facilitate the ubiquitous access and use of digital services.

This requires fully language-based interfaces, with unconstrained language input-output, keyboard-less operation, including multimodal interactivity, understanding of messages, interpretation of communicative acts and gestures.

- Providing **full multilinguality** for information content at all stages of the information life-cycle, and enable multilingual communications — business communication in particular.

This requires multilingual content generation and maintenance, multilingual authoring, content and software globalization and localisation, automated translation and interpretation, and computer-assisted language training.

The Structure

Specific HLT action lines will typically address more than one of the above challenges, and will consist of a mix of basic and applied research, demonstration, first-user validation and take-up, depending on the problem addressed. HLT activities will be broadly structured into three RTD strands:

April 1998

elsnet
•••••

- focused thematic clusters,
- transfer of technology to languages, and
- demonstration and take-up activities.

Each of these will cover their own needs in terms of RTD infrastructure. Accompanying measures, including standardisation and best practice, socio-economic impact assessment, dissemination and promotion of results, will complement RTD actions for improving their effectiveness and impact.

International cooperation, for which language technologies offer unique opportunities, will be an integral part of RTD activities. Special attention will be paid to multidisciplinary skills training, to serve the needs of a fast growing language industry.

Focused thematic clusters

HLT will implement a small number of thematic clusters of RTD activities addressing the global challenges referred to earlier. It will also provide a coordinated framework for — possibly comparative — activities, chosen for their relevance with respect to technology trends, market evolution, and societal developments. The aims of these clustered activities will be to

- provide a high degree of focus and better impact of RTD activities,
- improve the critical mass of European RTD and achieve a higher return on RTD investment,
- sustain the European leading edge in language-based digital services, and
- further a multidisciplinary approach.

High-priority themes will be identified periodically and selected in an open and flexible way, taking into account the rapidly changing socio-economic context.

Transfer of technology to languages

In order to close the time gap between the emergence of a leading-edge technology in one language and its availability as embedded technology for systems in other languages, HLT will support a timely transfer of key technologies to a broad range of languages within appropriate application frameworks, chosen for:

- the relevance and maturity of the technologies to be transferred and the pertinence of the chosen applications,
- the potential for impact on new business opportunities or expanding market penetration of established businesses,
- the contribution to levelling the ground for economic players and citizens across diverse linguistic and cultural backgrounds.

A strong industrial or socio-cultural drive, with the appropriate backing from the players involved, is expected for this type of transfer activity.

Demonstration and take-up

As in the past, HLT will be open to the needs of industry and society for solution-driven approaches, supporting a faster

uptake and broader market penetration of language technology applications and services. Support for sizeable 'bottom-up' demonstration activities with validation by early adopters, and first-use trials will be provided for services:

- in high added-value application domains,
- showing a strong user drive,
- providing convincing evidence of the technical, functional and economic viability of research results for real-life applications, and
- covering adequate life-size testbeds for new technologies.

R&D infrastructure

As already mentioned, RTD activities — projects and clusters — will cover their infrastructural needs, thus ensuring convergence of the two activity strands. This essentially concerns:

- **Evaluation** of language technologies and components to stimulate the pace of technological progress; and **benchmarking** of systems with respect to users (e.g. human factors, user acceptance) and businesses (e.g. processes, working conditions), to speed up technology transfer and facilitate decision-taking for early adopters. Together with impact assessment, these activities should also contribute to a better monitoring of HLT activities.
- **Language resources** for systems development in the context of RTD tasks, for on-line trade and trans-national services, for educational, commercial and corporate interactive publishing.

Common coordination and support will be provided for these activities as appropriate.

Next Steps

Before the launch of the first call for proposals, currently planned for the end of 1998 or the beginning of 1999, there will be further consultations with key actors in the field and with Member States. I am confident that with the continued cooperation of the community we will be able to provide a good general framework for HLT, and an appealing set of RTD activities to be called for in the near future.

FOR INFORMATION

Comments on and contributions to FP5 are warmly welcomed. Please send them by email to **Giovanni Varile** (giovanni.varile@lux.dg13.cec.be), mentioning 'HLT contribution'.

More information on FP5 can be found on <http://www.cordis.lu/fifth/src/news.htm>

The FP5 working document can be found on <http://www2.echo.lu/langeng/en/fp5/lt.html>

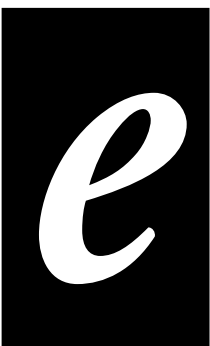
The report **Language Engineering: Progress and Prospects** (1997) can be ordered from:

LINGLINK Anite Systems

151 rue des Muguets

L-2167 Luxembourg

Email: linglink@anite-systems.lu



Making connections

New avenues for Spoken Language Understanding

Eva Hajicová and Petr Sgall, Charles University, Prague

*Speech Processing and Spoken Language Understanding were the focus of two conferences held in Greece last September: **Intonation: Theory, Models and Applications** (Athens) and **Eurospeech'97** (Rhodes). Eva Hajicová and Petr Sgall add some afterthoughts on the relationship between grammar and discourse, spoken and written language, and the relevance of all this for *Speech Processing and Natural Language Understanding*.*

Grammar and discourse

Classical (pre-Chomskyan and Chomskyan) linguistics, which focused on lexicon and grammar, did not achieve much in the domain of discourse, or dialogue in particular. Up to the 1970s, discourse-related issues were classified as belonging to 'parole' (or performance) and mostly considered beyond the horizon of research (with the exception of stylistics and a few occasional remarks).

Spoken Language Understanding, on the other hand, has to be based on an analysis of the whole process of communication, which includes aspects as diverse as discourse structure, the speaker's goals, strategies, attitudes, intentions and (objects of) attention, speech acts, conversational implicatures, and so on. And in the case of dialogue, all this has to be taken into account for a potentially large number of discourse participants. Obviously such a rich set of complex tasks cannot be approached without the only and limited tool that has been developed up to now for the analysis of language, namely the results gained in the study of grammar. The conceptual dichotomy which separates 'langue' from 'parole' is too simplistic in this respect.

Written vs spoken discourse

Another issue to be taken into account concerns the enormous differences between spoken and written language. Not only are these two types of discourse governed by their own norms; but texts are usually prearranged monologues, whereas spoken utterances are normally part of spontaneous dialogue. Written texts, which are typically linguistically regularized monologues (i.e. monologues which have been checked by their speakers for linguistically and cognitively relevant properties) are substantially easier for text-to-speech conversion or speech generation than spontaneous everyday dialogue. So it makes sense to tackle the former first; once progress has been made in this area, moving on to spontaneous 'unregulated' speech will become more realistic and productive. Alternatively the two domains could be tackled jointly in a mutually supportive way.

Results achieved so far in the areas of text linguistics and discourse analysis, which are limited but nevertheless significant, mostly concern written usage. But the irregularities and context-dependent phenomena identified in written texts are also relevant for speech, and it is crucial that this work is extended to the analysis of speech.

Connecting sentence and discourse

Returning to the possibilities offered by applications of grammatical analysis in the context of speech processing, it is important to consider these within an appropriate framework which takes into account the interactive nature of language. We believe such a framework should account for the topic-focus articulation of the sentence, and for the hierarchies of its underlying structure. The notions of topic and focus (of the sentence), of contextual 'boundedness' (as a grammatical counterpart of cognitive 'givenness') and of communicative dynamism (the underlying word order) may be helpful in connecting the sentence structure with the process of communication.

Current research on different kinds of sentential and phrasal stress (such as the work of **Pierrehumbert**, **Hirschberg**, **Beckmann**, **Steedman** or, in a slightly different context, **Selkirk**) moves in this direction. Steedman's recent work, in particular, supports our view that stress and other features of sentence prosody should be studied in relation to phenomena such as focus, topic and contrast. These phenomena have been described as part of the grammatical structures of languages; such grammatical frameworks now make it possible to formulate strong hypotheses regarding the position of different kinds of stress in utterances embedded in discourse.

Moving forward

We believe that research in the area of speech which connects these different aspects (and which has been rather rare up to now) may help overcome the difficulties hampering the exploration of issues such as

- how to account for the smoothness of a discourse as a sequence of utterances. If the relation between this smoothness and the well-formedness of a sentence is given due attention, then it becomes easier to explore issues at the level of the discourse which, with its fuzzy borderlines and various degrees and dimensions of acceptability, is by far more diversified than the sentence. The study of means such as the distribution of accents in different kinds of contexts is more effective if topic-focus articulation is taken into account as one of the fundamental aspects of sentence structure;
- how to overcome the scattered, 'parallel' character much research in linguistics in general now has, with researchers often being unaware of each other's work; and
- how to achieve an integrated, global description of language as a system, and its use in communication. Such a description is necessary for understanding speech and language as a single domain.

FOR INFORMATION

If you are interested in the line of research proposed in this article, please contact Eva Hajicová (hajicova@ufal.ms.mff.cuni.cz) or Petr Sgall (sgall@ufal.ms.mff.cuni.cz).

April 1998

elsnet
.....

Lessons learnt from the past, new paradigms for the future



Brian Oakley

Brian Oakley, chair of ELSNET's Industrial Panel, took part in the Telematics Reviews and was involved in setting up ELRA. In conversation with Mimo Caenepeel, he looks back on the Fourth Framework, and talks about changes and innovations he hopes to see in FP5.

ELNews: What are, in your view, the most important innovations, or strands, in FP5, particularly as it will affect Language Engineering?

Oakley: FP5 will differ from FP4 in that there will be much greater integration into fewer separate programmes and sub-programmes — or that at least is the intention. LE work could appear all over the place, but I have no doubt that in practice it will be in the “Creating a user-friendly Information Society” part of the programme. The total funds at the moment are somewhat down from the Fourth Framework, but my guess is that by the time they've finished arguing with the parliament it won't look very different.

From the point of view of LE, I suspect the most important innovation — if that is the right word — lies in trying to move back a bit from applicable work to prevent a gap opening up between basic research and applied work. The involvement of users in FP4 did result in a lot of the programme, particularly the Telematics part, becoming pretty user-oriented. I wouldn't say ‘applied’, but certainly closer to applications work than perhaps had been intended. Of course there was basic research in the long-range research programme. But research with a clear objective but no immediate applied endpoint, which is really quite important for LE, was not very well supported. Of course there are many aspects of LE which have been applied in all sorts of ways, and we sometimes forget they are the fruits of LE: things like spelling checkers and so on, which have become an everyday part of life. At the same time, when we look at things like Machine Translation or even, in my view, Speech Recognition, there is still a great deal of pretty basic work to be done before we get to the point where we

get the same sort of success as in so many other parts of IT. So it is really quite important that FP5 will try to provide more support for that gap. In particular people working at pretty basic applicable work will have the choice of either going to the long-range research programme, as now (and I think that programme will be at least as large in financial terms, and possibly slightly bigger, than it has been) OR — and this is perhaps the most important innovation — actually do that sort of work back within one of the main themes of the programme.

ELNews: You took part in the Telematics Reviews. What were the main points (suggestions, criticisms) that emerged from these evaluations? Do you think these have been taken on board in the new proposal? Do you feel there are areas that have not been sufficiently addressed — anything that isn't currently in the programme which you think ought to be in it?

Oakley: When we looked at the Telematics programme, we took the view that we needed more integration. The programme had proliferated, with ESPRIT, ACTS and the Telematics programme all dealing with applicable work, all having some LE in them for example. It really was time that it was integrated into a common whole. That was certainly taken on board in the proposal that went to the Commission. But if it were actually to be carried out in practice, it would only really be effective, to my mind, if there was a shake-up of the staff, with new people being allocated to the themes. It won't work if people remain in their fiefdoms and continue to divide and rule, as it were. It might be that such a shake-up will take place very soon, this Spring.

One of the big innovations in FP4 was undoubtedly getting the users involved. Everybody agrees that was a very good thing. But we hadn't expected that involving users would create a subtle shift in what they want. If you ask a user what R&D he wants, he will certainly tell you, but he will also tell you about policy issues which to him are comparable in impor-

tance, and which are perhaps delaying his investment much more than any lack of R&D. So I do hope that another innovation of FP5 will be that policy issues are taken more seriously. So that it would be perfectly possible for a research project to raise, for instance, web-related issues like ownership of property rights, confidentiality and censorship, which are not themselves new in the IT world, but which have definitely been brought into a different focus because of the vast expansion of the web.

One other major plank of our review — and here, unlike the others, I don't think there's much hope of the Commission being able to do much about it — was the question of bureaucracy. The fact of the matter is that the Commission is dealing with public funds, and that means that inevitably there has to be great caution and great care. I personally have a great deal of sympathy with the Commission over this. Nevertheless everybody — including staff in Luxembourg and Brussels — would like to see lighter bureaucracy. My main concern is the extraordinary limited amount of delegation of financial powers that goes on in the Commission. The powers delegated to the people that we deal with are really ridiculously low.

ELNews: What do you feel is the best way forward in terms of Resources and Evaluation? Do you feel it's useful in this respect to follow the American model (as happened with ELRA, and now ELSE), or do we need a radically different approach?

Oakley: Of course my last connection with the LE world was on the question of resources. I continue to believe that there is a major role for the Commission in that area, and I do think that the area will be taken more seriously in FP5. It's easy to forget how important things like standards and resources are, in all communication-related areas. If you build a highway you certainly need R&D, but you also need to invest in highway infrastructure and signalling; and you need international agreement on how you are going to put the roadsigns up. If you drive

e

from the EU into Switzerland they simply reverse the colours of the signs for motorways — it's infuriating... In LE too you have to have standards, you have to have infrastructure. We should not just expect the Commission to fund R&D, we should also expect them to help us with the resources and so on.

My view is that the Commission did a good job in setting up ELRA; the idea of trying to disseminate language resources more widely, of trying to create a market for them, is absolutely fine. And that's now underway; although it may not be an enormous trade, it's gradually building up. But I think there are much larger fish for the Commission to go for in that field. Let me give two examples of the

sorts of things which I'm afraid we may not see in FP5, but which I would very much like to see. First of all, taking language resources on the international scene. We've tried to deal with it within Europe, like the Americans have done within the USA. But at a time when we are trying to set up international exchanges and so on, it is quite ridiculous for Europeans to try, for example, to build up language resources for the Chinese languages, just as it's unreasonable to expect the Asiatic nations to invest heavily in language resources for the European languages. Why don't we try to get together on an international basis and do some large-scale deals, where the Chinese provide the Chinese language resources and we provide the European language resources and do some sort of commercial swap? That wouldn't necessarily be purely a barter arrangement; there's no reason why one shouldn't set up exchanges on a commercial basis as well. It would require the building up of confidence in the infrastructure, but I believe it could be done. Rather strangely I think we'd find that if we did do that, it would be the Far Eastern nations, and the Arab nations, that would be involved, rather than the Americans. Because although the Americans could join in, they'd be more likely to be duplicating resources stemming from the European languages.

The other thing that I'm interested in is taking seriously the development of language resources for lesser-used European languages. I believe that the maintenance of the languages of Europe is enormously important because the whole of our heritage and our culture depends on that language base. And the social consequenc-

es of things like not being able to communicate in your first language on the Web, are creating an even bigger gulf between the haves of the IT world and the have-nots. We should be taking major steps to prevent that gulf opening up. But it's unreasonable to expect the lesser economies to struggle with their language resources. It takes as much investment to provide language resources in a minority language as it does in a major language: the extent of the dictionaries and the corpora that are required are the same.

I would like to see the Commission taking this seriously. Not necessarily funding it all, but setting up a structure, rather like the old EUROTRA programme, where

“Why don't we try to get together on an international basis and do some large-scale deals, where the Chinese provide the Chinese language resources and we provide the European language resources, and do some sort of commercial swap?”

the Commission provided the glue and the organization, saw that the standards used were established centrally and maintained, but also worked with the individual nations and their governments and institutes to ensure that the necessary resources were provided. I don't know how the funding should go. But in some sense it's as important for the big language nations to have the smaller language resources as vice versa. Because after all, if they are going to trade with a minority language they need those language resources. So it's reasonable for the costs of this to be shared around the Union, not necessarily in proportion to the number of speakers of the various languages.

ELSNews: What do you think is the most useful role for networks in FP5?

Oakley: I happen to be a great fan of the ELSNET-type network organization. It seems to me that the achievement of the European research programmes has been above all a social one. At events like the recent Telematics conference in Barcelona, you see thousands, literally thousands of members of the scientific and technical community of Europe getting together. It's something which has made us automatically, naturally work with each other in Europe; it's broken down the national barriers easily. It's substituted for the old relationship where a graduate worker in Europe would go to the States for his

apprenticeship — now he or she probably goes to some other European country. That's a great step forward., but it does need fertilizing. And the networks are the best way of doing that.

I find that a publication like *ELSNews* is extremely useful for keeping me abreast of what is going on in the LE world of Europe, since I'm not so deeply involved as to know it all. And of course it creates a way of getting to know the other people, and what other work is going on.

I happen to be involved in setting up a new embryo network for Quantum Computing, a new paradigm in computing. Quantum Computing is based not on the classical laws (as in classical computing) but on the quantum-mechanical laws

of physics. And that enables certain things which, if not possible in classical computing, are extremely difficult to do. I think Quantum Computing is going to be important in the LE area because of the problem of looking into unstructured corpora. If you've got a structured database then there are algorithms which enable you to retrieve data within time proportional to the log of the size of the database. But if it's an unstructured database then the average time must relate to half the number of entries. Not so with Quantum Computing. A man called **Lov Grover** has produced an algorithm which demonstrates that with a quantum computer you could retrieve such unstructured information within time proportional to the square root of the size of the database, rather than linearly with it. That sort of improvement is really quite important in changing the way in which we tackle certain things.

So in this field we're setting up a network, creating a newsletter, bringing people together — something which I believe is an extremely important use of public funds by the Commission

FOR INFORMATION

Brian Oakley can be contacted at brian.oakley@pop-3.ukonline.co.uk. For recent work on Quantum Computing, see <http://www.cogsci.ed.ac.uk/~marke/NatComp/> One of the workshops at the forthcoming LREC conference in Granada will be devoted to **Language Resources for European Minority Languages** (see *Future Events* section, p11)

April 1998

elsnet
.....

Wonderland in sight?

Mimo Caenepeel, University of Edinburgh

Even the weather collaborated on 25-27 of March, when 65 ELSNET members gathered in an old seminary (and fresh conference centre) in Soesterberg (NL) to carve out a vision of Wonderland. The two conference days were overcast and chilly — perfect weather, as **Steven Krauwer** kindly pointed out, for keeping warm by contributing to workshop debate. But during the weekend that followed temperatures rose considerably — a welcome reward for those with APEX tickets.

Debate was at the heart of the Wonderland conference, which combined presentations and workshop sessions during the day with hearty discussions in the local cafe at night. What is ELSNET's Wonderland about, and how do we get there? In his opening speech, Steven Krauwer compared it to that other Wonderland, in which Alice found that things she would never have thought possible did exist; the Language Technology wonders we dream up now, Steven said, are only a fraction of what will be possible in a number of years' time. **Niels Ole Bernsen**, in another keynote talk, pointed out that the Wonderland scenario has all the ingredients of a drama, with lots of actors (who may or may not share the same interests), considerable perils and risks, a changing and unpredictable environment, and some potentially high stakes.

High on the Wonderland list of priorities is a showcase of exciting, attractive Language and Speech Technology that speaks to the user's imagination. What would be involved in putting on such a show, and how can the ELSNET community combine forces to make it happen? First and foremost, said Steven Krauwer, we need to exchange information on what is available within the ELSNET community, and work together to use, evaluate and promote it. Communication and collaboration — chief concepts behind the Wonderland idea.

But communication also means asking for clarification, and drawing attention to potential pitfalls. **Tony Hegarty**, of **Anite Systems**, outlined plans for setting up an electronic showcase under the **LingLink** project, coordinated by Anite Systems, to which ELSNET could contribute. The LingLink project is receiving considerable funding for making this happen. But does this not mean, asked Ole Bernsen, that ELSNET's seat is already occupied? Will the LingLink enterprise not sit on grassroots initiatives like that of ELSNET? This gave rise to other questions, and animated discussion. What role can ELSNET play, given that it has only very limited resources? What will motivate people to contribute? What kind of objects do we want to put on show, and what kind of audience are we trying to attract? And is the Web the appropriate medium for this?

Two hours into the conference, it was clear that there were plenty of issues to be addressed, and gaps to be bridged — including, as **Joana Lipeikiene** pointed out, the enormous difference between research conditions and resources in different parts of Europe. The format of the conference allowed plenty of time to discuss these, during six workshops on relevant themes, a special session on Eastern Europe, ample coffee breaks and a lavish Indonesian banquet.

What did ELSNET in Wonderland achieve? It put the ELSNET community in touch, gave people the opportunity to exchange ideas, information, research results and demos, and allowed them to voice their opinions, enthusiasms, biases and prejudices. When the event drew to a close, some people were clearer about what ELSNET's Wonderland might look like; others not. But everybody was clear about one thing: there is work to do.

We'll keep you informed.

The 6 main themes of the Wonderland conference

- *The Wonderland concept*
- *Training*
- *Multilinguality*
- *Information services*
- *Tools*
- *Resources*

For each of the themes there was a session with three or four presentations, followed by a brief discussion. In addition to this there were a number of parallel workshops on the same topics.

Topics addressed at the workshops

1. Identify NL/Speech products or services that are suited for inclusion in a permanent electronic exhibition (short term)
2. Identify NL/Speech products, services, facilities or prototypes that we could adopt for our own internal use (short term)
3. Formulate possible proposals for new sorts of NL/Speech products or services, to be submitted for funding from one of the EU or national funding bodies (medium term).
4. Formulate longer terms research actions on which the next generation of NL/Speech systems should build (longer term).

Proceedings of workshop results may be made available. For more information please contact ELSNET.

“Brain for brain, result for result, we are not one inch behind the Americans. But something else is missing.” (*Niels Ole Bernsen*)



Conference participants relax in the Cafe of conference centre Kontakt der Kontinenten.

e

Central and Eastern Europe

Central and Eastern European countries were well-represented at the conference. A special session on Central and Eastern Europe addressed questions such as

- How can we use our own technologies to bridge the linguistic, technological and other gaps between East and West?
- What perspectives does the extension of the EU towards the East offer, and how can we prepare ourselves for this?

ELSNET General Meeting

The conference closed with a general ELSNET meeting, which offered participants the chance to exchange views with the ELSNET Executive Board on ELSNET's activities in general. Topics discussed at the AGM included:

ELSNET Executive Board

The EB meets three times per year. ELSNET members are always invited to propose agenda points. Because of Ewan Klein's resignation, a new EB member has to be elected. Deadline for nominations was April 15, 1998.

Industrial Panel

ELSNET's Industrial Panel, which currently consists of seven representatives from industry, chaired by Brian Oakley (UK), will meet with the EB once a year. The next joint meeting will be in June 1998.

Annual Report

Every year an Annual Report will be sent to all ELSNET Members. Comments are welcome

Roadmap

The Commission has asked all ES-PRIT Networks of Excellence to provide an explicit time-dependent technological roadmap which shows the progress of the technology. An outline document of ELSNET's roadmap, with specific milestones will soon be sent to all ELSNET members for feedback.

Let's Twist Again

Marc Blasband, NS Utrecht

So where do we go next? Marc Blasband suggests a TWIST (Trying Wacky Ideas for Speech Technology). The HMM-model, he argues, has taken us as far as it goes. We need to break through to a new paradigm: Wacky Ideas are called for. Blasband sets out his vision for exploring such ideas while minimizing the risks they entail.

It is becoming clear that the HMM model is reaching the limit of its possibilities: the results, though impressive, are not sufficient, and a quantum leap is needed. But the size of the corpora required for the next step, quadrigrams, is frightening.

Searching for new directions, on the other hand, brings its own difficulties and risks. The current state and quality of HMM-based work is the result of enormous amounts of time and effort spent by many researchers worldwide, and is therefore hard to abandon. And, assuming that a new idea is found and pursued, there is always the risk that funds will dry up quickly in the absence of clear and rapid results.

Moreover, the chances of any one group being able to create something that can compete with the results of the years of work that went into HMM are very small. Such a group would run the risk that the idea itself turns out to be bad (but this only becomes apparent with hindsight); or that the first results are disappointing, especially in comparison to HMM-based results; or that significant details are missing.

A breakthrough is necessary, but I would argue that it can only be achieved through a big-scale collaborative venture, with a number of research projects starting at the same time and progressing in parallel. In the scenario I have in mind, each project would study a different idea, which would have to be 'daring and wacky' (otherwise there would be no point to the whole concept). As the projects advance, strong interactions between the groups would be organized to ensure optimal changes and amendments to the different streams of research. An objective validation procedure to measure progress would have to be agreed upon, and existing corpora would be used to build and validate the proposed implementations.

In this type of approach, all the participants would partake equally in (the advantages of) any positive results of any of the projects. At the same time, the risk of one idea not bringing the results hoped for would be reduced for everybody. Some of the wild ideas that could be considered:

- Go back to the rule-based approach that HMM replaced;
- Mix the rule-based approach and HMM;
- A new architecture with a more cyclic, less waterfall structure;
- Use elements of chaos theory as fractals;
- Use genetic algorithms;
- Finite state machines;
- Genetic algorithms
- Using complex perceptrons (more than three levels, with training back-tracking loops).

If enough researchers are interested in such a collaborative research programme, it could be started under the auspices of ELSNET, which should help in obtaining the necessary funds from the various funding agencies.

FOR INFORMATION

If you are interested in the ideas put forward here, please contact Marc Blasband (cplr@worldonline.nl).



Marc Blasband

"We don't have enough clean, reliable, available data at affordable prices." (Susan Armstrong)

April 1998

elsnet
.....

Minutes of the February 1998 ELSNET EB Meeting summarized

The first meeting in 1998 of the ELSNET Executive Board was held in Athens, on February 16, 1998. Many thanks to Professor George Carayannis and his staff at the Institute for Language and Speech Processing for their hospitality.

The following topics were addressed:

Executive Board vacancy

Ewan Klein (UK) has resigned from the ELSNET Executive Board. Ewan was one of the founders of ELSNET, and ELSNET coordinator from 1991 till the end of 1994. Because of his resignation, a new member for the ELSNET Executive Board has to be elected. The ELSNET Executive Board invites ELSNET nodes to propose candidates for



Ewan Klein

Executive Board membership. Out of the list of all proposed candidates the Board will appoint one member (for a period of two years), taking into account the profile and qualifications of the candidates, as well as the overall composition of the Board in terms of geographical distribution and thematic orientation.

ELSNET-ELSE Workshop

ELSNET and ELSE are organizing a joint workshop at the First International Conference on Language Resources and Evaluation in Granada. The workshop, *Towards a European Evaluation Infrastructure for NL and Speech*, will take place on Wednesday May 27 from 09:00 -13:00 hours. Prominent speakers from field have been invited to present papers addressing the implementation of such an evaluation scheme at an international scale, in terms of motivation, advantages, and problems it may give rise to. The workshop will finish with a one-hour panel discussion. Target communities are people involved in evaluation exercises (as organizers or participants), as well as those with an interest in the role of industry in evaluation.

LE Training Showcase

ELSNET is supporting the development and evaluation of six small Distance Learning tutorials in the area of Language and Speech, which will serve as pilots for the eventual development of an Internet LE Training Showcase. Careful monitoring and evaluation of the pilots will provide important guidelines for innovative learning and training opportunities in language and speech technologies. The work will be carried out in collaboration with the SOCRATES Thematic Networks **Advanced Computing in the Humanities** and **Speech Communication Sciences**.

ELSNET Bullet course

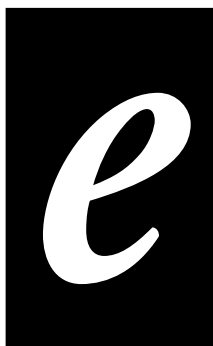
In September 1998 the **Humanities Information Technologies programme (UIB/HIT)** at the University of Bergen will organize an ELSNET Bullet course on *Terminology systems in Translational Information and Documentation Management*. The course will give an overview of current methods in electronic management of terminology, spanning a range of interdependent phases, including organizing documentation and capturing data from running text; storing, editing, maintaining and updating data using various data structures; and electronic publication. It will also discuss the achievements and shortcomings of current methods, applications and prospects for the future, in the framework of recent theory of Terminology. There will be on-line demonstrations of both the design and user possibilities of Terminology Management Systems, and participants will be able to get hands-on experience with a terminology management workbench.

Next meeting

The next ELSNET Executive Board meeting, a joint meeting with the ELSNET Industrial Panel, will take place in Utrecht, the Netherlands, on Monday June 22, 1998. ELSNET members are invited to propose agenda points before June 7.



Members of the EB at the Athens meeting. From left to right: Antonio Zampolli, Gerrit Bloothoof, Björn Granström (top), George Carayannis (bottom), Steven Krauwer, José Pardo, David Cornwell, Roberto Cencioni, Nikos Fakotakis, Joseph Mariani.



May 13-15, 1998: *Twendial'98: Formal Semantics and Pragmatics of Dialogue*, University of Twente, Enschede, Netherlands. Further info: Email: joris@cs.utwente.nl URL: <http://www.seti.cs.utwente.nl/Parlevink/Conferences/twlt13.html>

May 24, 1998: *Interaction Agents Workshop*, L'Aquila, Italy. Further info: Email: avi-ii@fub.it URL: <ftp://fub.it/pub/AVI-II98/>

May 25-27, 1998: *STiLL - ESCA-Workshop On Speech Technology In Language Learning*, Stockholm, Sweden. Further info: Email: still@speech.kth.se URL: <http://ophale.icp.grenet/escal>

May 26, 1998: *Adapting Lexical and Corpus Resources to Sublanguages and Applications*, LREC workshop, Granada, Spain. Further info: Email: velardi@dsi.uniroma1.it

May 26, 1998: *The Evaluation of Parsing*. LREC Workshop, Granada, Spain. Further info: Email: john.carroll@cogs.susx.ac.uk URL: <http://www.icp.inpg.fr/ELRA/confre.html>

May 27, 1998: *Language Resources for European Minority Languages*, LREC workshop, Granada, Spain. Further info: Email: briony@cstr.ed.ac.uk URL: <http://www.icp.inpg.fr/ELRA/confre.html>

May 27, 1998: *Towards a European Evaluation Infrastructure for NL and Speech* (LREC workshop jointly organised by ELSE and ELSNET), Granada, Spain. Further info: Email: steven.krauwer@let.ruu.nl URL: <http://www.icp.inpg.fr/ELRA/confre.html>

May 27, 1998: *Databases of Central and Eastern European Languages*, LREC workshop, Granada, Spain. Further info: Email: p.j.roach@reading.ac.uk URL: <http://www.linguistics.rdg.ac.uk/speechlab/research/babel>

May 28-30, 1998: *First International Conference on Language Resources and Evaluation* (LREC). Granada, Spain. Further info: Email: lrec@ilc.pi.cnr.it

May 28-30, 1998: *(Preferably) Non-Lexical Semantics*, Paris, France. Further info: Email: rz@ccr.jussieu.fr

May 30-Jun 1, 1998: *7th International Workshop on Nonmonotonic Reasoning*, Trento, Italy. Further info: Email: brewka@informatik.uni-leipzig.de URL: <http://saturn.hut.fi/~ini/nmrw98.html>

Jul 13-24, 1998: *Robustness: Real Life Applications in Language and Speech*. ELSNET's 6th European Summer School on Language and Speech Communication, Barcelona, Spain. Further info: Email: summer98@gps.tsc.upc.es URL: <http://gps-tsc.upc.es/veu/ess98/>

August 10-14, 1998: *COLING-ACL '98*, Montreal, Canada. Further info: Email: coling-acl98@iro.umontreal.ca URL: <http://coling-acl98.iro.umontreal.ca/MainPage.html>

September 25-29, 1998: *The future of the Humanities in the Digital Age*, Bergen, Norway. Further info: Email: futurehum@uib.no URL: <http://www.futurehum.uib.no/>

October 26 - 29, 1998: *Speech and Computer (SPECOM'98)*, St-Petersburg, Russia (Supported by ESCA and ELSNET). Further info: Email: specom@mail.iias.spb.su URL: <http://www.spiiras.nw.ru/speech>

Would you like to subscribe to *ELSNews*, or do you know someone who might be interested in ELSNET's activities?

To add someone to the *ELSNews* mailing list, please fill out the details below and return this form to the ELSNET secretariat (full address at the back of this issue).

Name _____

Affiliation _____

Postal address _____

Email _____

We sometimes make addresses available to organisations within the field of Language Engineering who provide information relevant to ELSNET members. If you do **not** want your address to be passed on to such organisations, please tick the following box.

April 1998



ELSNET Secretariat

Steven Krauwer
Coordinator

Mariken Broekhoven
Assistant Coordinator
Utrecht University (NL)

Task Group

Convenors

Training & Mobility
Gerrit Bloothoof, Utrecht University (NL)

Info Dissemination

Ewan Klein
Edinburgh University (UK)

Linguistic & Speech Resources

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

Research

Niels Ole Bernsen
Odense University and Joseph Mariani
LIMSI-CNRS

Industrial Panel

Harri Arnola, Kielikone (SF)
Roberto Billi, CSELT (I)
Michael Carey, Enigma (UK)
Jean-Pierre Chanod, Rank Xerox Research Centre (F)
Harald Höge, Siemens AG (D)
Bernard Normier, GSI-ERLI (F)
Brian Oakley (chair, UK)

ELSNET Participants Academic Sites

A OFAI/Univ. Vienna/Vienna Univ. of Technology
B University of Antwerp
B University of Leuven
BU Bulgarian Acad. of Sciences, Sofia
BY Belarussian Academy of Sciences, Minsk
CH IDSIA, Lugano
CH ISSCO, Geneva
CZ Charles University, Prague
D Univ. des Saarlandes, Saarbrücken
D DFKI, Saarbrücken
D IAI, Saarbrücken
D Univ. Hamburg
D Univ. Kiel
D Univ. of Stuttgart
D Ruhr-Univ. Bochum
D Univ. Erlangen
DK Ctr for Sprogteknologie, Copenhagen
DK Ctr for PersonKommunikation (CPK), Aalborg
DK Odense University
E Universidad de Granada
E Univ. Politecnica de Catalonia/Univ. Autonoma de Barcelona
E Univ. Politecnica de Madrid
E Univ. Politecnica de Valencia
F LIMSI-CNRS, Orsay
F IRIT, Toulouse
F Inst. de la Comm. Parlée, Grenoble
F IRISA, Rennes
F Laboratoire Parole et Langage-CNRS, Aix-en-Provence
F CRIN, Nancy
GR ILSP/NCSSR "Demokritos", Athens
GR Wire Communications Lab., Patras
H Hungarian Acad. of Sciences, Budapest
H Technical University, Budapest
I Ist. di Linguistica Computazionale, Pisa
I IRST, Trento
I Fondazione Ugo Bordoni, Rome

IRL University College Dublin
IRL University of Dublin
LT Institute of Mathematics and Informatics, Vilnius
N University of Trondheim
NL Stichting Spraaktechnologie, Utrecht
NL Inst. for Perception Research, Eindhoven
NL Leyden Univ.
NL Catholic Univ. of Nijmegen
NL TNO Human Factors Research Institute
NL Univ. of Amsterdam
NL Univ. of Tilburg
NL Univ. of Twente
NL Utrecht University (coordinator)
P INESC/ILTEC/Univ. Nova de Lisboa
PL Polish Academy of Sciences, Warsaw
RO Research Inst. for Informatics, Bucharest
RU Russian Academy of Sciences, Moscow
S KTH, Stockholm
S Univ. of Linköping
UK Defence Research Agency, Malvern
UK UMIST, Univ. of Manchester
UK Univ. of Cambridge
UK Univ. College London/School of Oriental and African Studies (SOAS)
UK University of Edinburgh
UK Univ. of Essex
UK Univ. of Dundee
UK Univ. of Leeds
UK Univ. of Sheffield
UK Univ. of Sunderland
UK Univ. of Sussex
UK Univ. of Ulster
UK Univ. of York

D Langenscheidt
D Novotech GmbH
D pc-plus Computing
D Philips Research Laboratories
D Siemens AG
D Verlag Moritz Diesterweg
DK Tele Denmark
E Telefonica I&D
F ACSYS
F Aerospatiale
F GSI-ERLI
F LINGA s.a.r.l.
F MemoData
F Rank Xerox Research Center
F Systran SA
F TGID
F VECSYS Speech Processing
GR Knowledge A.E.
H Morphologic
I CSELT
I Database Informatica
I Sogei (IRI-FINSIEL Group)
I Syntax Sistemi Software
I Tecnopolis CSATA Novus Ortus
I Olivetti Ricerca SpA
NL KPN Research Laboratories
NL Polydoc N.V.
RU Analit, Ltd.
RU Russicon Company
S Telia Promotor (Call Centre Division)
FIN Nokia Research Center
FIN Kielikone Ltd
UK ALPNET UK, Ltd
UK BICC plc
UK British Telecommunications
UK Cambridge Algorithmica Ltd.
UK Canon Research Centre Europe Ltd.
UK Enigma Ltd.
UK Hewlett-Packard Labs
UK Logica Cambridge Ltd.
UK Sharp Laboratories
UK SRI International
UK Vocalis Ltd.

Industrial Sites

B Lernout & Hauspie Speech Products
D aspect GmbH
D Daimler-Benz AG
D Electronic Publishing Partners GmbH
D Grundig Professional Electronics GmbH
D IBM Deutschland

What is ELSNET?

ELSNET, the European Network in Language and Speech, was established in 1991 with funding from ESPRIT Basic Research. There were 25 founding members of the network. Currently, there are more than 80 universities and research institutes, and more than 50 companies participating.

The long-term technological goal which unites the members of ELSNET is to build integrated multilingual NL and speech systems with unrestricted coverage of both spoken and written language. Building multilingual NL 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 both the language and speech communities. In this respect, it is important to note that a network like ELSNET can only function well if all members of the network are prepared to give and to receive.

Electronic Mailing List

elsnet-list is ELSNET's electronic mailing list. Email sent to elsnet-list@let.ruu.nl is received by all member site contact persons, as well as other persons who have an interest in ELSNET's activities. This mailing list may be used to announce activities, post job openings, or discuss issues which are relevant to people in the European natural language and speech communities. To request additions/deletions/changes of address in the mailing list, send mail to elsnet@let.ruu.nl.

ELSNET web pages

Detailed information about ELSNET and its activities and publications is available on the Web at the following URL: <http://www.elsnet.org>. Comments and suggestions for new web pages are very welcome.

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.ruu.nl
WWW: <http://www.elsnet.org>

