elsnews

9.4

HLT Products Win IST 2001 Prizes

*This year's European IST Prize list includes two companies whose products have direct relevance to the fields of Human Language Technologies. ELSNews extends its congratulations to **Sympalog**, from Germany, who have developed a toolkit for speech recognition, understanding, and dialogue, called **FullConversation**, and to **Auralog**, from France, for their comprehensive language learning course, **TeLL me More**.*

*ELSNews asked both companies to give us an idea of the research background of their products. **Bernard Muller** from Auralog, and **Manuela Boros** from Sympalog have kindly provided the information below. But first, for those readers who have never heard of the IST Prize ...*

What is the European IST Prize?

The European IST Prize, now in its seventh year, is an award for ground-breaking products that represent the best of European innovation in information technology. It is organised by the European Council of Applied Sciences and Engineering (Euro-CASE), with the sponsorship and support of the Information Society Technologies (IST) Programme of the European Commission.

The IST Prize provides public recognition to entrepreneurial teams that excel in generating novel ideas and R&D and converting them into marketable products. Twenty prize winners each receive EURO 5000, and an executive jury, composed of independent, highly respected Europeans, select three grand prize winners. Each grand prize winner receives EURO 200,000 and the European IST Prize Trophy. Whilst we congratulate the three grand prize winners, whose products were in the fields of health and business, we turn here to the HLT-related products **FullConversation** and **TeLL me More**.

Sympalog Speech Technologies AG

First thoughts of founding a company to develop and distribute high-end spoken dialogue systems emerged some three years ago, in 1998. At that time, a group of scientists at the Chair for Pattern Recognition of the University of Erlangen-Nürnberg and the Bavarian Research Center for Knowledge Based Systems (Forwiss) decided to transform their research in high-end spoken

dialogue systems into production systems.

The two research institutes, lead by Professor Niemann, are known for their ground-breaking work in speech recognition and dialogue systems. The founders of Sympalog decided to build a commercially focussed venture that would bring together a critical mass of speech systems expertise and deliver commercially viable systems. After significant groundwork, Sympalog was officially formed in March 2000.

Research prototypes prepared the way for the development of the Sympalog technology. Of particular note were those built for a number of European and German national projects, including SUNDIAL, SQEL, and Verbmobil. Verbmobil, for example, contributed valuable knowledge in the practical processing of prosodic information.

The goal of the company is to provide spoken dialogue systems that people find natural to use. Sympalog systems are much more flexible than existing telephone interaction systems, allowing the caller to choose freely what to say and when to say it. Sympalog products are modular, and may be used separately or combined with other vendor's modules, such as other recognisers.

Sympalog's **FullConversation** is a toolkit that allow the construction of information applications supporting fluent conversational voice interaction between people and computer. It consists of a number of Sympalog modules, ranging from speech recognition to speech understanding and dialogue.

The recogniser **SympaRec** is designed for speaker-independent recognition of telephone and cellular phone quality speech. It performs especially well for spontaneous, colloquial language. Other features, such as prosodic information, are extracted from the speech signal to enhance speech understanding and subsequent dialogue.

The application-independent dialogue engine can be easily configured to support new complex applications. It provides rapid turn-round of required changes, and totally eliminates the need to define a rigidly structured dialogue process.

Sympalog **InfoScanners** perform part of the speech understanding process by extracting context-relevant chunks of information, such as time, names of airports, stocks, or account numbers from utterances.

SmartPrompt allows the configuration of flexible application dialogues. It performs most of the core dialogue functionality, including database access and prompt generation.

Lessons Learnt / Future R&D Directions

On entering the market for spoken dialogue systems, it was clear that *natural dialogue* means very different things to different people, including Sympalog customers and partners. The success of natural dialogue, as opposed to menu-based systems, will depend on being able to demonstrate its significant advantages. For example, the automation of far more complex tasks will bring enormous gains in user friendliness and ease of use.

The desire of clients to implement more complex applications will push the requirements of advanced dialogue systems even further. Great efforts will be needed to manage this complexity, such as how to optimise feedback information to the user during a dialogue, in order to make subsequent interaction as easy and as quick as possible.

FOR INFORMATION

IST Prize: http://www.ist-prize.org

Sympalog has recently become an ELSNET node
Sympalog: http://www.sympalog.de

Auralog: http://www.auralog.com

Auralog

TeLL me More – The Solution is a comprehensive language learning course, with both off-line pedagogical support and an on-line tutoring service. It is a personalised tool that adapts to the learner's objectives and level. The activities include interactive dialogues, a glossary and a dictionary, grammar and conjugation rules, and over 1000 exercises. TeLL me More uses the latest in multimedia technologies.

Auralog initially specialised in speech recognition technologies. TeLL me More history started around 10 years ago. The first product launched on the market was called AURA-LANG, an MS-DOS software based on specific hardware (like dedicated DSP cards, as PCs were not powerful enough to handle speech recognition). It was the first time that speech recognition had been used in foreign language learning.

Auralog worked on a couple of European projects around 1994, but most of the research for TeLL me More was handled by Auralog's technical team itself. For the speech recognition, Auralog builds its own higher technological layers on the basic technologies from different sources (e.g., L&H, IBM, Dragon). This way Auralog can remain relatively independent of those technology providers, bring great added value from its own research, and still benefit from the improvement to those raw technologies.

With speech recognition technology, TeLL me More recognises what you say, assesses your pronunciation, and corrects any mistakes. It also highlights mispronounced words using the exclusive SETS technology (Spoken Error Tracking System), which is a typical example of Auralog-specific technologies.

Auralog has concentrated a lot of its research efforts into getting SETS to work in an acceptable manner, finally succeeding in 2000. Of course, there will still be work on improvements to SETS, aimed at producing more precise and in-depth feedback for the user.

Auralog's research is used to develop additional, complementary tools which are incorporated into new versions of TeLL me More, all the time aiming at improving the ease and efficiency of foreign language learning. To improve pronunciation, for instance, users can listen to the model, try to imitate it, and get feedback on their performance compared with that of the model. They can compare graphs of energy versus time, compare the pitch curves (fundamental frequency versus time), identify any mispronounced word with SETS and practice the pronunciation of any word alone (out of context). They can also learn how to produce each phoneme of the language by using TeLL me More's 3-D phonetic animation facility, which was developed in response to the need for a way to explain how each elementary sound of the language is pronounced. It allows users to watch a step-by-step breakdown of the articulation as they listen to it, with follow-up training on several words and sentences that include the mispronounced phoneme.

The software developed at Auralog has to adapt to users' needs, and a constant effort is made to think of, and design appropriate tools for all situations.

The Open Language Archives Community

Steven Bird, *University of Pennsylvania*
and Gary Simons, *SIL International*

Steven Bird

Fifty years from now, language researchers digging through the 'linguistic record' will observe a silent period. Looking back from their future perspective, we can see that this period lasted a decade or two, and extended from the time language researchers began using computers to store their data, up till the time they began using XML and Unicode. Much of the data created during this period has survived, but most of it is almost useless since we have no way to recover its structure or interpretation. A digital museum of the 2050s which documents the history of language resources may have a 'showroom of worst practice', devoted to turn-of-the-century language resources. Venturing into this room, an inquisitive language researcher of the future is drawn to an exhibit consisting of a tastefully lit pile of 3.25" floppy discs (well, okay, a digital image thereof). The nearby plaque reads:

Abroc Dictionary c. 2001

A dictionary of the extinct language Abroc. The material is thought to be in Word 5 format. Documentation for its 8-bit font is lost, leaving us to guess the interpretation of many code points. A dozen phrases from the Abroc language, thought to have originated from this dictionary, were important in an unresolved theoretical controversy dating back to the early 2030s.

Back in the present now, we see that the web and its associated technologies put us in a position with tremendous potential for creating, sharing, and re-using language resources, but also for chaos and frustration on an unprecedented scale.

Proposers of language resource projects regularly promise: "*we will make all results of this project available to the community over the World Wide Web*". In reality what often happens is that they force the data into relational form, store it in a proprietary database, and expose a restricted and poorly documented interface on the web. A couple of years later, when a central server is upgraded, the database stops working. The sole surviving artefact, rescued from the disc five years later when the machine is finally expired, is a binary database file – encrypted heritage, another dinosaur of the digital epoch.

It gets worse. In spite of the web, we have great trouble just *finding* the resources we want. The *Corpora* mailing list – and others like it – are dominated by queries about the existence and availability of a certain kind of resource. A selection of recent corpora subject lines read: "*Tagged Arabic corpora for grammar extraction?*", "*Tools needed to process British National Corpus*", "*References for Swedish taggers & parsers?*".

Posters seek advice about the suitability of particular resources for particular purposes, but it is often difficult to decide on a good course of action when the primary information is an unco-ordinated set of suggestions originating from strangers on a mailing list. *ELSNews* readers will be all too familiar with the difficulty of finding appropriate data, tools, and advice.

The Philadelphia Workshop

In December 2000, we organised a four-day workshop on Web-Based Language Documentation and Description, held at the University of Pennsylvania. The meeting brought together a group of some 100 linguists, archivists, software developers, publishers, and funding agencies, who are responsible for creating language resources in North America, South America, Europe, Africa, the Middle East, Asia, and Australia. Sponsorship was from the US National Science Foundation and the Institute for Research in Cognitive

Science. The goal of the workshop was as follows (from the workshop website):

This workshop will lay the foundation of an open, web-based infrastructure for collecting, storing, and disseminating the primary materials which document and describe human languages, including wordlists, lexicons, annotated signals, interlinear texts, paradigms, field notes, and linguistic descriptions, as well as the metadata which indexes and classifies these materials. The infrastructure will support the modelling, creation, archiving, and access of these materials, using centralised repositories of metadata, data, best practice guidelines, and open software tools.

The workshop itself consisted of about forty presentations, three panel sessions, and several working group sessions. On the first day, we presented our vision for an Open Language Archives Community built on the Open Archives Initiative. The remaining presentations covered metadata for describing language resources, the concerns of various stakeholders, descriptions of projects, and demonstrations of systems.

The Open Archives Initiative

The Open Archives Initiative (OAI) was launched in October 1999 to provide a common framework across electronic pre-print archives, and it has since been broadened to include digital repositories of scholarly materials, regardless of their type.

In the OAI infrastructure, each participating archive implements a repository, a network-accessible server offering public access to archive holdings. The primary object in an OAI-conformant repository is called an *item*, having a unique identifier and being associated with one or more metadata records. Each metadata record describes an archive holding, which is any kind of primary resource such as a document, data, software, a recording, a physical artefact, a digital surrogate, and so forth. Each metadata record will usually contain a reference to an entry point for the holding, such as a URL or a physical location.

To implement the OAI infrastructure, a participating archive must comply with two standards: the *OAI shared metadata set* (Dublin Core), which facilitates interoperability across all communities participating in the OAI, and the *OAI metadata harvesting protocol*, which allows software services to query a repository using HTTP requests. These OAI 'data providers' typically have a submission procedure, a long-term storage system, and a mechanism permitting users to obtain materials from the archive. An OAI 'service provider' is a third party that provides end-user services, such as search functions over union catalogues, based on metadata harvested from one or more OAI data providers. The following figure illustrates a single service provider accessing three data providers. End-users only interact with service providers.
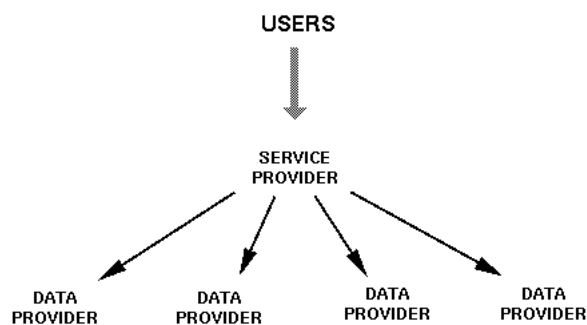
A Service Provider Accessing Multiple Data Providers

The Open Language Archives Community (OLAC)

Workshop participants resolved to begin construction of OAI-compliant language archives and an OLAC metadata set that would permit focussed searching for the resources held in those archives. The OLAC metadata set will extend the Dublin Core set only to the minimum degree needed to express what is fundamental about open language archiving. For example, the *subject language* of a resource should be specified with a language identifier (such as provided by ISO-639), not a text string. The Ethnologue 3-letter language codes will extend coverage from some 450 languages to over 6,800, initially by using the extension method permitted by RFC-1766 (see the paper by Simons in the online proceedings). OLAC will need other controlled vocabularies for such things as openness, platform, language resource type, other resources that are needed to use the item, and so forth.

Linguist List has agreed to harvest and index the OLAC metadata from all language repositories, and offer a centralised union catalogue for all language resources. We anticipate that other organisations will offer more specialised service providers. To date, some 16 archives

are participating, and we already have OAI-compliant data providers for half of them. LDC has a demonstration service provider that indexes some of these data providers.

Beyond cataloguing resources, OLAC would like to promote best practices for storing primary materials, to help end the 'silent period' we described at the beginning. Workshop participants shared a strong suspicion of rigid requirements, yet a willingness to adopt practices voluntarily once their usefulness has been demonstrated. To this end, the OLAC Gateway will maintain a collection of Requests For Comment (RFCs) and recommended best practices. RFCs may report an existing practice and experience with that practice, then present a case for wider adoption of the practice. RFCs may be accompanied by other materials, such as an XML DTD.

OLAC has strong European involvement with its Advisory Board and Alpha Testers. We seek more European involvement, particularly from many language archives which are not yet involved.

Websites:

OLAC Gateway, including prototype data providers and an experimental service provider:
http://www.language-archives.org/

Proceedings of the Workshop on Web-Based Language Documentation and Description
http://www.ldc.upenn.edu/exploration/expl2000/

Open Archives Initiative:
http://www.openarchives.org

Language Archives Page:
http://www.ldc.upenn.edu/exploration/archives.html

Linguistic Data Consortium (LDC):
http://www.ldc.upenn.edu/

Linguist List:
http://www.ling.ed.ac.uk/linguist/

FOR INFORMATION

Steven Bird is Associate Director of the Linguistic Data Consortium, and Adjunct Associate Professor of Computer Science and Linguistics at the University of Pennsylvania. Steven is a principal investigator on three NSF projects (TalkBank, ISLE, and Linguistic Exploration). He co-edits the Cambridge Series in Natural Language Processing, and has recently been elected to the executive committee of the Association for Computational Linguistics. Before coming to Pennsylvania, Steven was based at Edinburgh University for 11 years, and undertook research on computational phonology involving extended fieldwork in Cameroon.

Email: sb@unagi.cis.upenn.edu
Web: http://www.ldc.upenn.edu/sb/

Gary Simons was Director of Academic Computing for SIL International from 1984 to 1999, and directed several projects that developed software to assist field linguists in documenting and describing language, including IT, CELLAR, LinguaLinks, and FieldWorks. In his current position as Associate VP for Academic Affairs, Gary oversees this area, as well as SIL's efforts to launch an on-line language archive. During the development of the Text Encoding Initiative's guidelines for text markup, Gary was involved as a member of the Committee on Text Analysis and Interpretation, and of the Technical Review Committee.

Email: gary_simons@sil.org
Web: http://www.sil.org/sil/roster/simons.htm

Special Issue of *ELSNews* on Minority Languages

The next issue of *ELSNews* (10.1, Spring 2001) will be a special edition dedicated to work in the areas of NL and speech technology concerned with minority languages.

Contributions related to work in these areas are invited, to be received by the deadline of 17 April. Submissions for publication may be, for example, progress reports, announcements of resources (accompanied by some description), comments on existing resources, or other relevant material.

Copy of a more general nature is still invited, but may be held over to the following issue (Summer 2001).

The Changing World of Lexicography: Interview with Sue Atkins

Sue Atkins was General Editor on the Collins-Robert English-French dictionaries between 1967 and 1989, and Lexicographic Adviser to Oxford-university Press from 1989 to 1994. She is a Founder Member and Past President of EURALEX, was the initiator of the British National Corpus, and works with Charles Fillmore on the Frame Semantics paradigm. Sue was awarded an Honorary Doctorate by the UK's University of Brighton in August 2000, for her contribution to lexicography.

Jenny Norris *from* ELSNews *asked her about her career.*

ELSNews: *You began working as a professional lexicographer in 1966, when lexicography was done manually. What have been the greatest changes to the way you work and the tools you use since then?*

SA: Well, obviously, a tremendous change has come with the advent of computers. When I started in lexicography we were just given a word list for a particular letter and told to get on with it. So I set myself up with a handful of dictionaries and my 'H' list, and start working on all the different senses that I could think of for each word. We used boxes of filing cards, and for a complex word like *head* I would have about 70-80 cards. It took about 6 months of evenings to complete my 'H' list of English-French entries, relying heavily on my French lodger at the time. When I had finished my list, I posted it off to the publishers, and prayed that it didn't get lost in the post. Sometimes, after I had sent my list back to the publisher, I needed to check my entry for a particular word (if, for example, I had a similar issue to deal with and couldn't remember exactly what I had done before). The only solution here was to phone up the publishers and ask them to read out the relevant entry to me. So even things like photocopying and email have made an enormous difference.

Nowadays, with electronic corpora available, it's obviously very different, with, typically, 60,000 corpus lines for a word like speak and its various forms. Of course, you can't read, or – more importantly – retain, 60,000 lines, so the problem now is not getting access to enough information about a word, but knowing how to deal with all the information at one's disposal. So, for example, Adam Kilgarriff's work on developing tools for extracting information from a corpus in a way that doesn't require you to read all 60,000 lines for a word is highly important for lexicography.

As far as tools are concerned, at least we now know what we need: we need linguists to provide information for the lexical tools; lexicographers to inform linguists about what is required; and very sensitive computer scientists who will write programs that satisfy themselves for elegance and satisfy us for information.

But apart from the computer-related changes, the most noticeable change is in training. When I started with Collins as a lexicographer, there was no training – I was just taken on and told to get on with it in whatever way seemed best to me, as a bit of individuality was a good thing. By about three years later I was telling them that they really did need notes on how to do the job, especially since they were then employing other people and I was being asked to check over their work. That's the way that people learned how to write dictionaries – in teams, with somebody else reading what you had written and telling you what not to do! They were the worst of teaching methods. So I sat down and wrote a set of instructions – which turned out be 170 typescript pages – and that was the first comprehensive style guide.

Nowadays, certainly in many British publishing houses, there is good formal training for new lexicographers, and that is one of the things I'm currently involved in developing. In fact, we have our first training workshop in lexicography and lexical computing this summer, which is being held at the Information Technology Research Institute (ITRI) at the University of Brighton.

ELSNews: *How do you use a corpus to write a dictionary entry, and how much of what you know about a word gets put into it?*

SA: Basically, writing dictionary definitions is an analyis-based task: you look at the facts and then try to make the dictionary entry reflect them. You start by looking at the corpus lines and try to identify patterns for the word you require. The things that stands out most clearly are the subcategorisation frames that the word operates in, and these are very easy to find in a corpus, because you can sort on the right and then on the left context of the target word. So, if it's a noun, you can see, for example, all the adjectives that are modifying it. Then, of course, you find all the fixed phrases that the word gets used in. And then you have to see, much more subtly, how people use the word, and you have to collect together a lot of prototypical examples – and some way-out examples – of what you feel is the same sense, to allow you to write a definition (if it's a monolingual dictionary) which covers more or less all the prototypical examples and is broad enough to cover the way-out examples. The mistake that most people make is to choose the wrong genus term, and if you do that there's no way you will be able to write a good definition.

The other thing about writing a dictionary is that it's like a dialogue between you and the user: if the user is really trying and gets it wrong, then it's your fault and not the user's fault. It's important to remember who your users are: for example, a family dictionary can be used by anyone between the ages of about 10 and 90, and you must always ask yourself how much of a definition could be understood by someone in that age range who has no specialist linguistic knowledge. Then you look at all the information you have gathered, and you're lucky if you can get 10% of these facts into an entry.

In fact, I think it's a very bad entry if you put everything you know into a definition. Before the days of corpora, that was quite easy to do, by looking at a few other dictionaries and adding your own knowledge, but that can be misleading because, of course, you don't have all the knowledge about a word at your disposal, even with a corpus, let alone without one.

ELSNews: In your opinion, what are the biggest problems that lexicography has to face?

SA: The basic problem is that the lexical querying tools are not sophisticated enough to speed up the lexicographer's role in dictionary-writing, to make it an economic proposition to write a dictionary entirely from a corpus. So the only way forward is to improve the tools, and the only way to do that is to get lexical semanticists and syntacticians to inform these tools.

Otherwise, the main problem is lack of funding: dictionaries have to pay their way in a commercial market. Tight budgets don't allow you to make improvements which are now, thanks to the computer, technically possible; for instance, removing circular definitions (the bain of the dictionary critic). So, for example, a *ceremony* is a type of ritual, and you look up *ritual*, to be told it's a type of ceremony. I actually have more faith in corpora than dictionaries, and certainly, if I wanted to know the difference between *ritual* and *ceremony*, I'd look at the way the two words are used in a corpus. The facts are there, and longer definitions could make the differences clear, but to be fair, if you put that sort of detail into a family dictionary it would run to 20 volumes (and never sell).

The lack of funds is a problem for training as well: for publishers who think almost anyone can write dictionary entries, training can seem a waste of money.

Building and supporting lexicographically useful (I won't say 'balanced') corpora and smart corpus-query tools is a huge expense for even a large publisher, and it's difficult to persuade their accountants that it is worth it when their dictionaries already sell well. To build the British National Corpus we had to persuade publishers to accept the idea of 'precompetitive resources'; so far, they don't see lexical tools in this way, and commercial lexicography is the poorer for that.

ELSNews: Can you tell us a bit about what you're doing now and any plans you have for the future?

SA: At the moment my heart lies with the FrameNet project, which was started at the International Computer Science Institute in Berkeley, California, under Charles Fillmore. In lexicography, you need a theory of lexical semantics which links syntax with semantics, and is flexible enough to be expanded and changed, whilst still remaining coherent and consistent when faced with new data.

Fillmore's theory of frame semantics and construction grammar does this: it is flexible enough to deal with the problems that lexicographers have, and it's also powerful enough to give us a varying statement of lexicographical relevance, allowing us to explain which part of any context is relevant to the definition of a word.

In FrameNet I, we had humans annotating the semantic roles, the grammatical functions, and the phrasal analysis. One of our aims in FrameNet II is to automate as much of the manual tagging as we can. And an exciting development is EuroFrameNet, which is just being proposed by its member sites in Barcelona, Stuttgart, Erfurt, Paris, Pisa, and Brighton. The idea is to build European-language FrameNets, and the ITRI will be particularly involved in the formalisation of the language using Gerald Gazdar's DATR.

I'm also working on the EC ISLE project, focussing on the semantic component of lexical entries, and taking forward the work done in previous EAGLES and SIMPLE projects.

The other thing I'm very involved in at the moment is the Lexicom training workshop we're starting this year at Brighton. It's the only course of its type in the world – a one-week hands-on course that focuses on how to write dictionaries, and how to make best use of the computational tools available. We will also be offering an MSc in lexicography and lexical computing at the ITRI, starting next year.

Winter 2000-01

elsnet

SIGdial: Supporting Meetings on Discourse and Dialogue

David Traum, *University of Southern California*

David Traum

For our second SIGdial contribution to *ELSNews*, we summarise SIGdial's efforts in encouraging and publicising meetings to enable discourse and dialogue researchers to meet, exchange information, and collaborate. SIGdial's relations to meetings are at three levels, which are discussed below in order of increasing SIGdial involvement.

At the loosest level, SIGdial serves as an information collection point for meeting announcements and calls for participation. The primary vehicle is SIGdial's calendar page (currently http://www.pitt.edu/~dialcal/calendar.html, and always accessible from the calendar button on SIGdial's main page), which includes listings of upcoming events that may be of interest to SIGdial members. These events include both large conferences (e.g., ACL, COLING, CogSci) at which discourse and dialogue papers may be found, and workshops in related areas (e.g., SIGIR, ICoS), as well as specialised workshops focussing exclusively on (aspects of) discourse and dialogue. The calendar is frequently updated by SIGdial's information officer, and comprises two main sections, one for upcoming calls for papers (CFPs), and one for upcoming meetings. There is also an archival section, with links to web pages of previously held meetings, making the calendar pages a useful research tool.

SIGdial's second level of meeting support is a set of endorsed events. Organisers of the events apply to SIGdial's executive, according to specified guidelines (http://www.pitt.edu/~dialcal/endorsement.html), and the executive deliberates on endorsement, deciding whether the event has purposes and content central to the SIGdial charter, is likely to be of sufficient quality, and is open to SIGdial members. Organisers of endorsed events can advertise the endorsement both to potential funders and to the general public. There is also a separate page of endorsed workshops on the SIGdial site (http://www.pitt.edu/~dialcal/events.html), and endorsed events are indicated with the SIGdial logo in the main calendar. Endorsed events include serial events on discourse and dialogue topics (e.g., the workshop series on Semantics and Pragmatics in Dialogue, and that on Knowledge and Reasoning in Practical Dialogue Systems), as well as one-off workshops (e.g., the recent NAACL/ANLP workshop on Conversational Systems; the upcoming ESSLLI workshop on Information Structure, Discourse Structure and

Discourse Semantics, Helsinki, 20-24 Aug; and the ISCA Tutorial and Research workshop on Prosody in Speech Recognition and Understanding, New Jersey, 22-24 Oct.).

Whilst the endorsed events (and other meetings on the calendar) are a means of presenting new research results and high-quality interactions among discourse and dialogue researchers, there has been a perceived need for a more regular forum for the range of work in the field, neither too diluted by work outside the discourse and dialogue area, nor focussed only on a sub-topic, and moreover a place to come back with further results or follow-up to work presented in one-off specialty workshops. For these reasons, SIGdial has instituted its own workshop series, which is SIGdial's highest level of meeting support.

The first in this series was held in Hong Kong, in October 2000, just after the ACL meeting. This was a two-day workshop, with 20 presentations, and over 40 participants, including sessions on annotation and corpus work, discourse, dialogue management, and tools and evaluation of dialogue systems (for more information, including online proceedings, see http://www.sigdial.org/sigdialworkshop). Participants engaged in lively discussions following the presentations, and there was an overwhelming call for a second workshop in 2001. Plans for this are now underway, and the workshop will be held on 1-2 September 2001, in Aalborg, Denmark, just before the Eurospeech Conference (see the SIGdial website for a pointer to a CFP with full details). As well as being a venue for presenting individual papers, we hope that the SIGdial workshops can be a place for organising working groups and theme sessions on topical areas. Indeed, break-time discussions in the first workshop have already led to ideas for several task groups.

Letters to the Editor

Regular readers of ELSNews *will remember that our previous issue,* ELSNews *9.3: Autumn 2000, contained the last in a series of thought-provoking 'Opinion' columns by Professor Yorick Wilks. In reference to this piece,* ELSNews *has received three letters, which we print in full below. One of the letters is from a person in the field who is highly respected, but who has chosen to remain anonymous (name supplied).*

Readers who missed ELSNews *9.3 can access it via the ELSNET website at http://www.elsnet.org/. Yorick's opinion column is on page 9.*

To the Editor,

A light style like that used by Yorick Wilks in his last opinion column (*ELSNews* 9.3) is often a clever defence against serious criticism, since the critic risks appearing uncool in taking the target so seriously. Well, I have long lost any hopes of being cool, what with living as a non-native speaker with native-speaker children who correct my accent and word usage, learning how to drive and ski as an adult, and generally spending far too much time with code and calculations and not nearly enough exploring nice old-fashioned cities with Michelin-rated restaurants. So, I will be terminally uncool now and say bluntly what I think of Yorick's last opinion piece: he stopped one piece too late.

Under the cover of humorous similes like that between the ICCL and the House of Lords, Yorick's attack on the ACL – for that is what the piece really is – rehearses tired rhetoric against increased participation of women in science, against scientific progress, and against democratic institutions. Since I grew up under Salazar's paternalistic, philistine dictatorship, I am pretty familiar with both the style and the substance of such tirades, so I will just focus on the most appalling of the insinuations, that somehow gender has a causal role in ACL publication.

Yorick seems to have forgotten that ACL reviewing has been blind since 1993 (except for the Maryland ACL). He most certainly confuses correlation with causation – maybe the statistical *fashion* he derides would have come in handy after all. He does not consider the great differences among countries in female participation in science, which necessarily affect the proportion of women accepted as authors between conferences with different geographical coverage. Nor does he consider the fact that the proportion of women computer scientists has been increasing – all too slowly – in the West, which means that, relatively speaking, the women are younger, and are thus more likely to work on newer areas, which, again relatively, may be more strongly represented – as Yorick says – in ACL meetings.

To its credit, the ACL has positively welcomed a new generation of women computational linguists, who are adding greatly to the field. It is sad that such a senior

member of the field as Yorick shows total insensitivity to the obstacles to the scientific advancement of women, and is so ready to encourage prejudice and division within the field by making an accusation for which he has no factual evidence.

Yours sincerely,
Fernando Pereira

FOR INFORMATION

Fernando Pereira is Distinguished Research Scientist at WhizBang! Labs, the incoming chair of Computer and Information Science at the University of Pennsylvania, and a past president of the ACL.

Email: fpereira@flipdog.com

To the Editor,

Occasionally I am really shocked when a colleague I respect for creative research over a long career uses tactics I would have expected from a malicious stranger against members of our field.

When I read Yorick Wilks' attack on the change in the ACL's conference mission, I was astonished at the extensive hearsay and veiled attacks on other members of our shared field of study. Furthermore, I could not understand why it was necessary to defend COLING in such a way.

COLING and ACL have co-existed side by side for a very long time, and COLING's global goal has, to my mind, taught ACL the importance of holding conferences that are not bounded by a particular country. It might well be that there cannot be two international conferences in the field of computational linguistics, but I doubt it. As I see it, each conference has a different mission, and happily there is room for both, at least for the foreseeable future.

Among the attacks and hearsay in Professor Wilks' article is a snide remark about the need for sex equality in selection of papers. As a woman scientist in the field, I am amazed that Professor Wilks is uncaring about the

Letters (contd.)

bias that happens, even among scientists, towards those of a less represented gender, social class, less prestigious university, or less developed country.

Blind reviewing is one way to counteract some of this bias. It's not even all that hard to do, and ACL is only one of many conferences in the computer sciences (and the sciences as a whole) that operates with blind reviews.

Professor Wilks hints at the need for cooperation between COLING and ACL in the planning of meetings. I wish he had simply asked for this as a needed goal, going into the future, between the two organisations. Since he did not, I am doing so now.

Even unpleasant experiences can have good effects. While the opinion piece by Professor Wilks saddened me, it made me aware of the *ELSNews* publication, and I am delighted to have this newsletter to keep up on human language technology in Europe and elsewhere.

Yours sincerely,
Candy Sidner

FOR INFORMATION

Candy Sidner is Senior Research Scientist at Mitsubishi Electric Research Labs (MERL), past ("way past!") president of the ACL, and currently a reviewer for both the ACL and COLING.

Email: sidner@merl.com
Web: http://www.merl.com

To the Editor,

In his last article as columnist for *ELSNews*, Yorick Wilks gives some excellent reasons why ICCL (the committee that organises COLING conferences) and ACL should work together, or perhaps even merge:

- ICCL has no formal organisation, chapters, or membership, while ACL does have these
- ICCL has no bank account, which ACL does have
- ACL is becoming more international
- ACL conferences are increasingly resembling COLING, in terms of size and range of papers
- ACL already handles the sales of past COLING proceedings.

But then, in an astounding feat of logic, he concludes exactly the contrary! The suppleness of Yorick's reasoning never ceases to amaze. I will miss his column.

Yours sincerely,
Anonymous Fan
(name supplied)

We showed these comments to Yorick, who had the following to say:

To the Editor,

The point of Fernando Pereira's that is absolutely correct is that my tenses slipped in the sentence where I referred to improving one's chances of acceptance in an ACL conference by changing one's first name to a woman's (and correspondingly changing your country for COLING). I know indeed that ACL has had anonymised reviewing since 1993, as has COLING since at least 2000. However, and after apologising for the slip, can I add that I thought it obvious that it was a joke and that no one could imagine I meant it seriously! Before that time there had indeed been positive discrimination for women's papers in ACL to some degree, as anyone who has sat though the annual committee meetings can testify. Positive discrimination was then thought a respectable posture to adopt; it is less so now, of course.

In case anyone reads this abject apology who has not also read my original article, can I point out that the complaints have been about a single sentence, right at the end of a thousand words or so. I therefore find the invocation of Salazar's fascist dictatorship a little overheated, not to mention the suggestion that, in making that joke, I am in some way opposed to, or uncaring about, women's rights, scientific progress, and democracy.

Indeed I am not, as anyone who knows my politics and personal track record also knows. What I am sure of is that positive discrimination does not pay in the long run, and science is not always a democratic business. I certainly had no intention of offending (or shocking!) distinguished US ACL colleagues; of course not, for the article was for Europe, where a sense of irony is slightly more developed. On the other hand, the bit about the complexities of intersociety cooperation in my article was ironic, but definitely not a joke, since when writing I had just come from a meeting where a distinguished ACL figure had proposed that COLING should, in effect, close down! I am sorry, too, if more of that meeting seeped into my prose than I had intended.

Yours sincerely,
Yorick Wilks

FOR INFORMATION

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

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

Pubs, Poets, and the Weather: CL and NLP in the Dublin Area

Josef van Genabith, Dublin City University

In the past, Dublin and Ireland have attracted visitors for a number of reasons, prominent among them its pubs and poets. Over the last decade the country has undergone a substantial transformation – so much so, that some visitors (and locals) find it difficult to reconcile the old with the new Ireland. The economy is booming. The country enjoys near full employment. There is talk of a 'Celtic Tiger'. Ireland has become a net immigration country attracting professionals from Europe, the USA, and Asia. To a large extent, the economic success story is due to an Information Technology industry with software development, software localisation and globalisation, and support services. Business is attracted by favourable corporate taxes, Irish membership in the Euro zone, an English speaking workforce, and the availability of qualified and well-educated graduates.

Dublin City University in winter

Because of the high concentration of IT industry it is no surprise that Ireland, and the Dublin area in particular, provides an attractive setting for computing and CL/NLP degrees and research. The Dublin area alone features four universities: Dublin City University (DCU), University College Dublin (UCD), Trinity College Dublin (TCD) and the National University of Ireland in Maynooth (NUIM). What is more, the city hosts a large number of further third level technical and professional colleges. To add to this, the MIT Media Lab Europe research centre has recently decided to set up in Dublin.

All four Dublin universities offer degrees in computing, and three of the four, DCU, UCD, and TCD, have a keen interest in NLP. Indeed, two of those, DCU and TCD, offer popular and thriving undergraduate degree courses in CL/NLP: the BSc in Applied Computational

Linguistics (ACL) at DCU, and the BA in Computer Science, Language, and Linguistics (CSLL) at TCD. Both degrees are firmly rooted in computing and include a healthy dose of foreign language and computational linguistics. In addition to this, UCD has recently started a BA in Computer Science programme that allows students to combine any arts subject (such as, for example, linguistics or psychology) with computing. Graduates of these programmes are in great demand in both the Irish and international IT industry.

Apart from employment opportunities, Dublin has lots to offer to CL/NLP and computing graduates: TCD runs a one year taught MPhil course in Speech and NLP; UCD offers an MSc in Cognitive Science programme. In addition to the taught programmes, DCU, TCD and UCD all offer well established MSc-by-research programmes.

Graduate programmes take us firmly into research. The Dublin universities offer PhD programmes in computing, CL, and NLP, often with competitive scholarships and funding for qualified students. Due to the revenue generated by the thriving economy, for the first time in its history Ireland is developing an attractive infrastructure to support research in a substantial way.

CL/NLP is well placed in the research landscape. We are fortunate to have established a research community spanning different universities and departments in Dublin, this being largely due to the joint Dublin Computational Linguistics Research Seminars (DCLRS). The basic idea is simple and effective: rather than running separate (and often competing) departmental seminars, we pool our resources: DCLRS is hosted and funded by TCD, UCD, and DCU and rotates between the participating universities on an annual basis. This spreads the costs of an international research seminar series and, most importantly, is community-building: it provides an opportunity for staff, students, and industrialists working on NLP to meet on a regular basis. In deep respect to Irish culture, tradition, and hospitality, often these meetings are continued in the above-mentioned pubs long into the night. A similar situation holds for a weekly CL/NLP reading group. The difference is that unlike the research seminars, the reading group meetings both start and end in a pub!

To give you an overview of the CL/NLP research carried out in the Dublin area, below I cluster researchers around (very) broad research topics (rather than around departments and institutions). I cannot hope to be fully comprehensive here, so please use the names and

references as pointers to more in-depth information. To facilitate this, readers are invited to consult http://www.compapp.dcu.ie/~josef/dublin.html, which provides links to the institutions and people listed.

In the area of speech, Dublin is fortunate to have a very active speech, acoustic phonetics, and computational phonology/morphology scene, including Julie Berndsen-Carson (UCD), Ailbhe N' Chasaide (TCD), Christer Gobl (TCD), John McKenna (DCU), and Ronan Scaife (DCU).

In the fields of Syntax and Semantics, research is carried out in LFG, HPSG, GB, and CG. Research in semantics covers formal and computational, as well as cognitive/psycholinguistic and AI-oriented approaches. The researchers involved in syntax and semantics include Carl Vogel (TCD), Tim Fernando (TCD), Martin Emms (TCD), John Saeed (TCD), Arthur Cater (UCD), Mark Keane (UCD), Cathal Doherty (UCD), Andy Way (DCU), Sean O'Nuallain (DCU), Fintan Costello (DCU), and Josef van Genabith (DCU).

As far as corpus and data-oriented approaches are concerned, there is research into the development and maintenance of mono- and bilingual corpora, the use of corpora in human translation, automatic annotation of corpora and exploitation of corpus resources in robust and chunk parsing, and statistics-based machine translation. Researchers include Martin Emms (TCD), Jennifer Pearson (DCU), Dorothy Kenny (DCU), Andy Way (DCU), and Josef van Genabith (DCU).

Dublin is a thriving centre in the field of Computer Assisted Language Learning (CALL) research, with lively groups interested in both theoretical and applied CALL. Key researchers here include David Little (TCD), Christine Appel (DCU), and Françoise Blin (DCU).

There is some work which applies neural network approaches to NLP, and these are explored mainly by a group centred around Ronan Reiley (UCD).

We also have some thriving research on the interface between NLP and Multimedia, Information Retrieval and Extraction. This ranges from the development of state-of-the-art engineering applications to cognitive science oriented research, with Alan Smeaton (DCU) and Sean O'Nuallain (DCU).
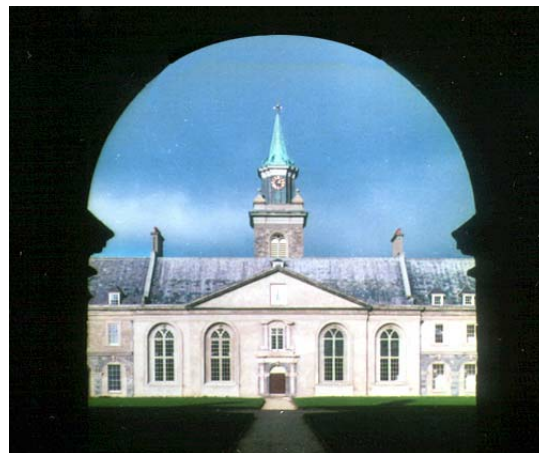
The Royal Hospital Kilmainham, now the Irish Museum of Modern Art in Dublin

Dublin is a very attractive place to learn about CL/NLP by taking one of its excellent undergraduate degrees. It is an exciting place to do research as part of a postgraduate degree, or as a future colleague at one of the universities or research centres. It is a fun place to live and work, to visit, or to give a talk, and I would be very happy to welcome *ELSNews* readers to Dublin. Is all well and rosy then? According to some statistics, the annual rainfall in Dublin is supposed to be less than that of Nice ... well, now that is one thing I find hard to believe!

FOR INFORMATION

Josef van Genabith is senior lecturer in the Computer Applications department of the Faculty of Computing and Mathematical Sciences at Dublin City University.

Email: josef@compapp.dcu.ie
Web: http://www.compapp.dcu.ie/~josef/

An HTML version with live links to everybody mentioned in the article is available at http://www.compapp.dcu.ie/~josef/dublin.html

The European Student Journal of Language and Speech WEB–SLS (Sponsored by ISCA, EACL and ELSNET)
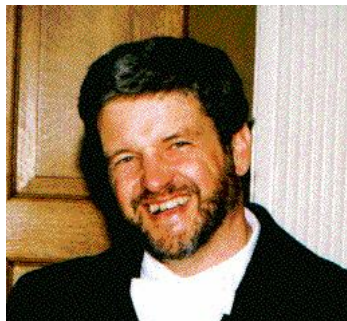
WEB-SLS is pleased to announce the extension of its policy to include summaries of on-going PhD theses. This will enable students working in neighbouring topics to be informed on one-another's work and exchange ideas, experience, and initiate future co-operation. From now on, in addition to publishing extended PhD summaries (accepted upon supervisor confirmation only) and regular papers (subject to favourable review), WEB–SLS will also accommodate summaries (up to three pages) of on-going PhDs that explain the main aspects of the research

Contributions should be sent to the Editor-in-Chief, Nick Fakotakis (**fakotaki@wcl.ee.upatras.gr**)
For other information, please see the WEB-SLS site at **http://www.essex.ac.uk/web-sls/**

Fence Sitting

John Nerbonne, *University of Groningen*

John Nerbonne

We extend a warm welcome to our new columnist, Professor John Nerbonne, from the Rijksuniversiteit Groningen in the Netherlands. Professor Nerbonne will be sharing some of his views with us over the next few issues.

We in the *ELSNews* community study language computationally, both with an eye to understanding language and especially language processing better, and with an eye to improving existing products and services involving language processing. Asked whether we're involved in science or engineering, a lot of would prefer to say *both!* and emulate illustrious predecessors in many branches of science (and engineering!) who combined work on pure science and application. (Or is it applied science and research engineering?)

Einstein and Szilard, not content with leading work in theoretical physics, also obtained patents for innovative refrigeration pumps. Kepler not only systematised Tycho Brahe's observations into his famous laws (rotational period proportional to radius, equal area swept in equal time), but likewise applied his work on equations for ellipses to the problem of measuring the contents of beer barrels (in Stuttgart!). Closer to our own field, neither Turing nor von Neumann were shy about tackling practical problems.

Examples of engineers with scientific credentials don't spring to mind as quickly, but maybe that's because scientists get more press to begin with, and because engineers often cultivate a 'just folks' public image. Edison worked hard on his *simple man* image, with his remark about genius being *1% inspiration and 99% perspiration*, but he corresponded extensively with Maxwell, and knew all the relevant physics in his areas of work.

There is a brief, but deservedly respected tradition in which research projects ask serious scientific questions while building demonstration vehicles that the same work enables. Woods et al.'s LUNAR system developed the ATN as a processing model even while showing it off in question-answering. The blackboard architecture developed in the HEARSAY project was put to immediate use in improving speech understanding. Verbmobil combined deep and shallow processing, and was likewise ambitious in supporting novel telephone translation.

But it won't do to recall that good science and engineering can go hand in hand. Sometimes they don't harmonise well at all: Pat Hayes has a nice piece in which he recounts how the then president of the US Academy of Sciences, Simon Newcomb, scoffed at the prospect of human flight – even after the Wright brothers' early successes! And even when science and engineering do co-exist in a single mind, they don't necessarily co-exist in any given piece of work springing from it. Einstein didn't make relativistic refrigeration pumps – he was just clever in more than one way.

Proposals and evaluations from computational linguistics are often treated with more then the usual refereeing savagery by funding agencies who see us waffling about science versus engineering, where we see ourselves as cleverly combining them. The savagery is not entirely misplaced. To begin, projects with modest resources do not find it easy to innovate in several ways simultaneously. HEARSAY and Verbmobil are not typical.

And we've all seen mediocre projects with explicit practical objectives gradually transform (by final report time) into theoretical *studies*. You get the feeling that if the project had only decided definitely – either for theory or for application – then the effort would have been more successful, and in any case clearer. And even practical success can be confusing in mixed-mode projects. A project several years ago aimed to support language learning with a focus on applying a particular grammar theory and formalism. The researchers tried out a prototype and, yes, people who used the system learned faster. But they were compared to others who learned from books based on traditional grammar. One group used interactive drills, the other a book. One group was tracked automatically, the other had to be self-paced. What was the gain in scientific knowledge? And what should we apply? ... There's always the follow-up project.

FOR INFORMATION

John Nerbonne is Professor in the Department of Humanities and Computing at the University of Groningen in the Netherlands.

Email: J.Nerbonne@let.rug.nl

Web: http://www.let.rug.nl/~nerbonne/

The First Meeting on Speech Technology – 1MST 2000 – in Spain

*Luis Fernando Rodríguez Romero, University of Seville*

The First Meeting on Speech Technology took place from 6 to 10 November, 2000 in Seville, in the south of Spain. The event was held at the *Tobacco Factory* (which now belongs to the University of Seville), a XVIIth century building, and setting of Bizet's masterpiece *Carmen* – a romantic environment for the first national conference on Speech Technology to be held in the country. The conference programme included all areas related to speech technology: speech recognition, speech synthesis, dialogue systems, machine translation, language processing, speech perception, language analysis, and language generation. The programme included thirty-one main presentations, seven invited talks, and posters. Most of the participants and contributors were from Spain, although other European and non-European countries were also represented: Great Britain, Ireland, Italy, and Canada.

The old Tobacco Factory, venue for the event

The 1MST 2000 was born out of a local initiative conceived by Dr López-Soto, who wanted to offer the University of Seville the opportunity to get in touch with speech technology advances in Spain. The original idea was to organise a seminar or course for the benefit of local researchers and students. This seminar would cover the main areas of speech technology: speech recognition, dialogue systems, and speech synthesis. The invited researchers were Luis Villarrubia (Telefónica I&D), Dr Hernández Gómez (Engineering School, Madrid) and Dr Rubio Ayuso (Engineering School, Granada). However, the seminar attracted the attention of many other research centres in the country, and the idea of organising a more general meeting was born.

The event was finally designed in the form of a national meeting or symposium for researchers across the country, and a call for papers was issued. The main incentive
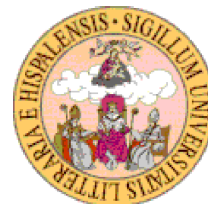
behind the meeting was the need to establish a regular (annual) conference in Spain, and to strengthen the co-operation and links between the different research groups in the country. To accomplish this aim, a special meeting was held on 8 November. More than thirty representatives from different universities and research centres in Spain participated, together with representatives from different companies located in the country: Telefónica I&D, Philips Speech Processing, IBM, Infospeech, Natvox, and Clic. A promising outline proposal to create a National Network on Speech Technology was devised, with the participation of all the aforementioned companies plus fourteen different Spanish universities.

One of the most recurrent strands of work presented at the meeting was the importance of linguistic-based methods in the field of speech recognition. An interesting presentation was made on the application of parsing methods to obtain N-best candidates from a DAG. Linguistic knowledge has not yet been fully incorporated into speech recognition techniques, although some nice attempts were reported at the meeting, particularly in the processing of prosodic features. However, the general approach was a probabilistic one. A recurrent statement was the need to develop systems that incorporate more linguistic insight.

Most of the research presented was in Spanish, although the three other official languages – Catalan, Basque, and Galician – were also widely represented. The companies present at the meeting (some of whom gave presentations) all agreed on the need to intensify the application of Speech Technology to these languages, especially in the area of telecommunications and dialogue systems. One of the highlights of the meeting was the presentation of some interesting investigations carried out at Universidad Autónoma de Madrid, which attempted to incorporate dialogue technology into a robotic environment. The University of Seville is currently involved in a robotic project funded by the European Fifth Framework, and the Universidad Politécnica de Catalunya is also working in the field of integrating speech recognition into robotics. Other important research groups in the country presented their latest investigations: the University of Vigo and the

University of Granada in the field of speech recognition, the Universitat Politécnica de Valéncia in the field of machine translation (I was particularly attracted by the application of neural nets into a machine translation prototype developed in Valencia), the University of the Basque Country in the field of speech synthesis, the Universidad Politécnica de Madrid in the area of speaker identification and dialogue systems, and the University of Barcelona and the University of Seville in the field of natural language processing.

The success of the event and the quality of the presentations guarantee a fairly open communication stream between all the participants, research centres, and companies, and it is very likely that a second meeting will be organised this year.

FOR INFORMATION

Luis Fernando Rodríguez Romero is a Research Fellow at University of Seville

Email: deutaro@yahoo.com

For more information on the meeting see http://fing.cica.es/TecnoHabla/index.html

A copy of the proceedings can be obtained from local organisers: teresa@fing.cica.es

---

Opportunities Under the New eContent Programme

eContent Announcement

A new, market-oriented *eContent* programme was launched by the European Commission in January 2001, with a budget of EURO 100m over four years. It aims to stimulate development and use of European digital content on the global networks.

For further information see www.cordis.lu/econtent, or contact econtent@cec.eu.int.

Specially relevant to the ELSNET domain is the second of three Action Lines, *Enhancing content production in a multilingual and multicultural environment*. For Action Line 2 and language technologies in general, see www.hltcentral.org, or contact hlt@cec.eu.int.

Project Proposals

The first calls will be published in March and November; early information will be available on the above websites.

Projects and other collaborative actions are expected to address three broad communities:

•private and public e-content players aiming to enhance their offerings through cost-effective internationalisation and localisation strategies

• businesses and public-sector agents intending to strengthen their e-commerce presence through, e.g., web marketing, retailing, and customer-care offerings adapted to diverse linguistic and cultural user requirements

•private/public partnerships geared towards wider deployment and exploitation of public-sector information.

Evaluator/Reviewers

The Commission invites suitable experts to register as evaluator/reviewers for eContent proposals and projects (New registrations are required for members of existing European panels of experts.)

First-call evaluation is scheduled for early July.

Online registration is available at www.cordis.lu/econtent/expert_form.htm; for information and assistance, contact infso-experts.econtent@cec.eu.int.

Job Openings

There will shortly be job openings in a dynamic team at Unit INFSO/D4 (Linguistic Applications of the Information Society), managing sizeable research and non-research projects within the eContent and related programmes. Successful applicants will be offered a 12-month contract as A-grade auxiliary agent working at the INFSO premises in Luxemburg.

Candidates should be graduates with at least five years relevant professional experience, preferably in a business environment or a primary R&D lab. They should have excellent spoken and written English, and very good analytical and communication skills. Previous experience with transnational environments and multi-party projects would be an asset.

For further information contact hlt@cec.eu.int; applicants should submit a two-page CV with recent photograph and relevant references to that address by 31 March 2001.

Winter 2000-01

elsnet

Future Events in 2001

March 18-21 *Human Language Technology Conference (HLT 2001)*: San Diego, California, USA.
Email: allan@cs.umass.edu;   URL: http://hlt2001.org

March 19-30 *Vilem Mathesius Lecture Series 15*: Prague, Czech Republic.
Email: cmejrek@ufal.ms.mff.cuni.cz; URL: http://ufal.mff.cuni.cz

March 23-25 *3rd North American Symposium on Corpus Linguistics and Language Teaching*: Boston, USA.
Email: corpconf@umb.edu

March 29 *Corpus Linguistics 2001 Workshop*: Lancaster, UK.
Email: amyi@cogsci.ed.ac.uk;   URL: http://www.ltg.ed.ac.uk/%7Ejeanc/corpus-linguistics

March 29-30 *International Colloquium on Trends in Special Language and Language Technology*: Brussels, Belgium.
Email: rita.temmerman@ehb.be;   URL: http://ttk.ehb.be

Mar 30-April 2 *CORPUS LINGUISTICS 2001*: Lancaster, UK.
Email: mcenery@comp.lancs.ac.uk ;   URL: http://www.complancs.ac.uk/ucrel/cl2000.html

April 2-3 *Workshop on Innovation in Speech Processing (WISP 2001)*: Stratford-upon-Avon, UK.
Email: sjc@sys.uea.ac.uk;   URL: http://www.sys.uea.ac.uk/wisp-2001/

April 26-28 *1st International Workshop on Generative Approaches to the Lexicon (GL2001)*: Geneva, Switzerland.
Email: Pierrette.Bouillon@issco.unige.ch; URL: http://issco-www.unige.ch/gl2001.html

April 27-29 *33rd Poznan Linguistic Meeting on Challenges for linguistics in the 21st century*: Poznan, Poland.
Email: plm@ifa.amu.edu.pl; URL: http://elex.amu.edu.pl/ifa/

May 30-June 4 *DIALOGUE 2001*: Moscow, Russia.
Email: info@dialog-21.ru; URL: http://www.dialog-21.ru/English/default.htm

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

Note that the list of ELSNET member nodes, which usually appears in this space, has been held over to make room for the late-breaking eContent announcement on page 15.

What is ELSNET?

ELSNET, the European Network of Excellence in Human Language Technologies, is funded by the European Commission's Human Language Technologies programme. Members are academic and public research institutes (83) and industrial companies (55) from all over Europe.

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

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

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

Electronic Mailing List

elsnet-list is ELSNET's electronic mailing list. Email sent to elsnet-list@elsnet.org 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@elsnet.org

Subscriptions

Anyone wishing to subscribe to *ELSNews* may do so online, by visiting **http://www.elsnet.org**, and following the links to *ELSNews* and subscription.