CLARIN

Newsletter

Can Research Infrastructures Bring A New Renaissance for European Research?

John Wood
Imperial College, London
Chair of the European Research Area Board

It is amazing that I can still remember some key developments that took place in ESFRI meetings over the years as if they were today. The most memorable was the endorsement of the first roadmap for research infrastructures where every member state present voted to accept the document. Just a few minutes before on the 28th September there had been an objection to one sentence and after a time out it was resolved at 17.58 hours precisely. Ah yes I remember it well!! As they say, since then, the rest is history. However my mind goes back further than that to when I first proposed the creation of a roadmaps for Europe. I had been made aware and was envious of the roadmap that the Department of Energy had produced in the USA and I could see that we needed to have something similar. I guess, at the time I had little idea of what it would finally look like. The decision to go ahead was taken in a crowded room in the Dutch Academy of Sciences in Amsterdam. The major difference between the US one and the one we agreed to produce for Europe was that it should cover all disciplines including arts and humanities. We agreed to form three working groups, one in arts and humanities, one in bio-medical sciences and one in physical sciences and engineering. Over lunchtime I let it be known that I would quite like to chair the one in arts and humanities. My naïve reason at the time was that I love literature

and music and I thought it would be more fun. However under the persuasive chairmanship of Hans Chang he asked me firmly to chair the physical sciences and engineering if the delegates agreed. They did.

The role of ESFRI

Just arriving at this stage had been a struggle. Having been a member of the initial working group that set up ESFRI there had, from the start been a tension about the powers of ESFRI and the role of the Commission. The Member States had made it very clear that they did not want the Commission to use ESFRI to make decisions on Research Infrastructures but to remain as a forum for exchange of views. What authority would the roadmap have given this background? Also, were Member States ready to take part in the exercise? Fortunately the Council subsequently endorsed and requested the need for a roadmap. To this day I am never sure how this happened or whether the Council understood the consequences of what they were releasing. However the Commission asked for proposals and this caused some confusion in the community at large. Around 200 very immature proposals were received but none, at that time had the support of the funding agencies in the Member States. From this chaos ESFRI developed a stage-gate process for judging the maturity of proposals and whether they were truly research infrastructure supported by Member States . Issues such as the need for open access, continuing research programmes dominated those early discussions. In the middle of all this the new Member States joined and their delegates were under considerable political pressure at home to "win" projects for themselves. By this time I had become the chair of ESFRI and I was delighted to see how the ESFRI delegates from the newer Member States matured and accepted their responsibilitiy for delivering solutions for the whole of Europe.

As members of CLARIN will know the project made it onto the first edition of the roadmap in 2006. Days after publication I was contacted by several other countries including China, Japan and the USA to come and discuss the roadmap and their co-operation in developments. One of the projects that was constantly referred to was CLARIN and this has continued ever since on

visits to South America and elsewhere. Just why were Government Agencies so interested in what, to some, might appear an obscure project? It was clear that they all saw the possibility that it would contribute to international agreements. Is there a possibility that international treaties and trade could be optimised if we had a language tool that would be truly un-ambiguous in any context and at any time. Now I suspect most members of CLARIN will not recognise this aspect of the project, but I am afraid there is a political reality that some of us have to face up to. Whether CLARIN can truly be a vehicle for world peace, only time will tell, but the expectations are high. Now if you are working on some aspects of the language of eleventh century philosophy you might gag at the expectations being placed on you. All I would say, is stick to your part of the project and leave it to others to look at the greater impact if that makes you comfortable.

CLARIN supporting cultural diversity

Having said that I personally would love to know more about the individual elements of real research as the project matures. I am fortunate to be on the Advisory Board of the British Library, and one of the thrills for me has been the analysis of texts that can be done whereby corrections made in ink on ancient manuscripts can be traced back to the monastery or wherever they were done and linked to the prevailing teaching of the time. We need to ensure that in the present age which tends towards mono-culturalism that we do not accept the lowest common denominator, but build on our diversity. I see CLARIN as a key element in supporting this action. So as you battle away with the possibility of an ERIC and all the other things necessary for a vibrant Research Infrastructure I do believe we are changing for ever the way research is undertaken and history will look back at this time as a pivotal period in Europe's development.

Since ceasing to be chair of ESFRI I have now become chair of the European Research Area Board and we have just published our vision for research in Europe entitled "A new Renaissance for European Research." Can I encourage you to look at it and see if you agree with our vision. We certainly see Research Infrastructures as one of the key elements in this vision. C

Editors' Foreword

**Marko Tadić
& Dan Cristea**
CLARIN Newsletter editors

Welcome, dear readers, to the issue of CLARIN Newsletter that closes the second year of our preparatory phase. Looking back to the initial, preliminary and preparatory meetings from three years ago it would be hard to imagine that our idea of research infrastructure for humanities and social sciences will start to shape up so soon. CLARIN started to attract interest from other related fields, but also from the direction of top level scientists who are involved in strategic planning and directing the science policy in Europe today. One of such figures is certainly John Wood, chair of the ERA Board who made a contribution to our front page reminding us of the tedious and not always straightforward process of discussing and accepting RI proposals that passed ESFRI filtering.

Other contributions in this issue also reflect the stage of the project we are in at the moment: the planning is well advanced, the first Startup Federation of Service Providers is being built and Kirster Lindér has report on that recent development.

The monumental digitisation project *Monumenta Linguae Dacoromanorum: The 1688 Bible Project* is presented by Gabriela Haja and Emil Munteanu. This project clearly shows how LR&T in humanities could provide the necessary shift in research paradigme thus leading to the e-Humanities in the proper sense.

This time the central part of the Newsletter is reserved for NEERI Manifesto which clearly requests that copyrighted material should be freely available for research purpuses in the Information Society and Knowledge Economy which we are both facing.

RI without knowledgeable users serves no purpose, so CLARINers soon realised that another project dealing with training users and maintainers of CLARIN RI is needed. The start of this project named CLARA is described in contribution by Koenraad de Smedt.

Our traditional report on national CLARIN and LR&T situation this time comes from Hungary and from the keyboard of Tamás Váradi.

Enjoy your reading. **C**

Call for contributions

Dear readers of the CLARIN Newsletter,
If you have ideas, thoughts, comments, additions, corrections, arguments, questions etc. which are connected to the CLARIN project, even remotely, please feel free to send them to us as your contribution at newsletter@clarin.eu or directly to the editors at marko.tadic@ffzg.hr and dcristea@info.uaic.ro.

List of national correspondents

Austria
Gerhard Budin

Belgium – Flanders
Inneke Schuurman

Bulgaria
Svetla Koeva

Croatia
Marko Tadić

Czech Republic
Karel Pala

Denmark
Bente Maegaard
Hanne Fersøe

ELRA/ELDA
Stelios Piperidis
Khalid Choukri

Estonia
Tiit Roosmaa

Finland
Kimmo Koskenniemi

France
William Del Mancino
Bertrand Gaiffe

Germany
Lothar Lemnitzer

Greece
Maria Gavrilidou

Hungary
Tamás Váradi

Italy
Valeria Quochi

Latvia
Andrejs Vasiljevs

Malta
Mike Rosner

Netherlands
Peter Wittenburg

Norway
Koenraad De Smedt

Poland
Maciej Piasecki

Portugal
Antonio Branco

Romania
Dan Cristea
Dan Tufiş

Spain
Nuria Bel

Sweden
Sven Strömqvist

UK
Martin Wynne

CLARIN calendar of events

Here is a list of CLARIN events and events from the fields of language resources and language tools that may be of interest to CLARIN members.

December 2009

2009-12-04 to 2009-12-05: TLT8: The Eighth International Workshop on Treebanks and Linguistic Theories, Milan, Italy

2009-12-07 to 2009-12-09: eScience 2009: 5th IEEE International Conference on e-Science, Oxford, UK

January 2010

2010-01-18 to 2010-01-20: ICGL 2010: The Second International Conference on Global Interoperability for Language Resources, Hong Kong, China

February 2010

2010-01-31 to 2010-02-04: GWC 2010: The Fifth Global Wordnet Conference, Mumbai, India

2010-02-05: CLIN 2010: Computational Linguistics in the Netherlands, Utrecht, The Netherlands

March 2010

2010-03-21 to 2010-03-27: CICLing 2010: Conference on Intelligent Text Processing and Computational Linguistics, Iaşi, Romania

2010-03-25 to 2009-03-26: POS Tagging 2010: POS Tagging Conference, Mysore, India **C**

CLARIN builds the Startup Federation of Service Providers

Krister Lindér
University of Helsinki

In the beginning of 2010, five CLAR-IN centres in Germany, Finland and the Netherlands will form an initial start-up Service Provider Federation which will be used to test all procedures and which, once tested, will be extended quickly to more centres in more countries. Already this startup federation will provide more than 1,000,000 researchers and students in the Humanities in the three countries with improved access to online tools and resources based on trust principles. The resources available for research purposes in one country can now be made available to researchers in the other countries via single sign-on. Single sign-on means that each researcher can use his or her own university user name and password as a unique identifier recognized by all the service providers and it is the basis, for example, to build virtual collections of resources from different centres.

Current state of federation

The federation currently consists of the CSC-IT Center for Science in Finland, the Berlin-Brandenburgische Akademie der Wissenschaften/DWDS and the Institut für Deutsche Sprache (IDS) in Germany, and the Institute for Dutch Lexicology (INL) and the Max Planck Institute for Psycholinguistics (MPI) in the Netherlands. All service providers have agreed that the Max Planck Institute will sign agreements with the national identity federations on their behalf. When the CLARIN European Research Infrastructure Consorttium (ERIC) has been established then this role will go over to the ERIC.

The Service Provider Federation has made agreements with national identity federations, i.e. HAKA in Finland, SURFnet in the Netherlands and DFN in Germany, to be able to register as a service provider for universities and research institutions. It is the task of these institutions to provide identity information about the researchers during the authentication process, so that the researchers can seek permission online to access the research tools and resources of the Service Provider Federation.

It needs to be stressed that the right to define who is permitted to access their resources remains with the individual repositories, i.e. the agreements are not intended to assign any rights or obligations of any one Service Provider to another or to the Service Provider Federation. This includes intellectual
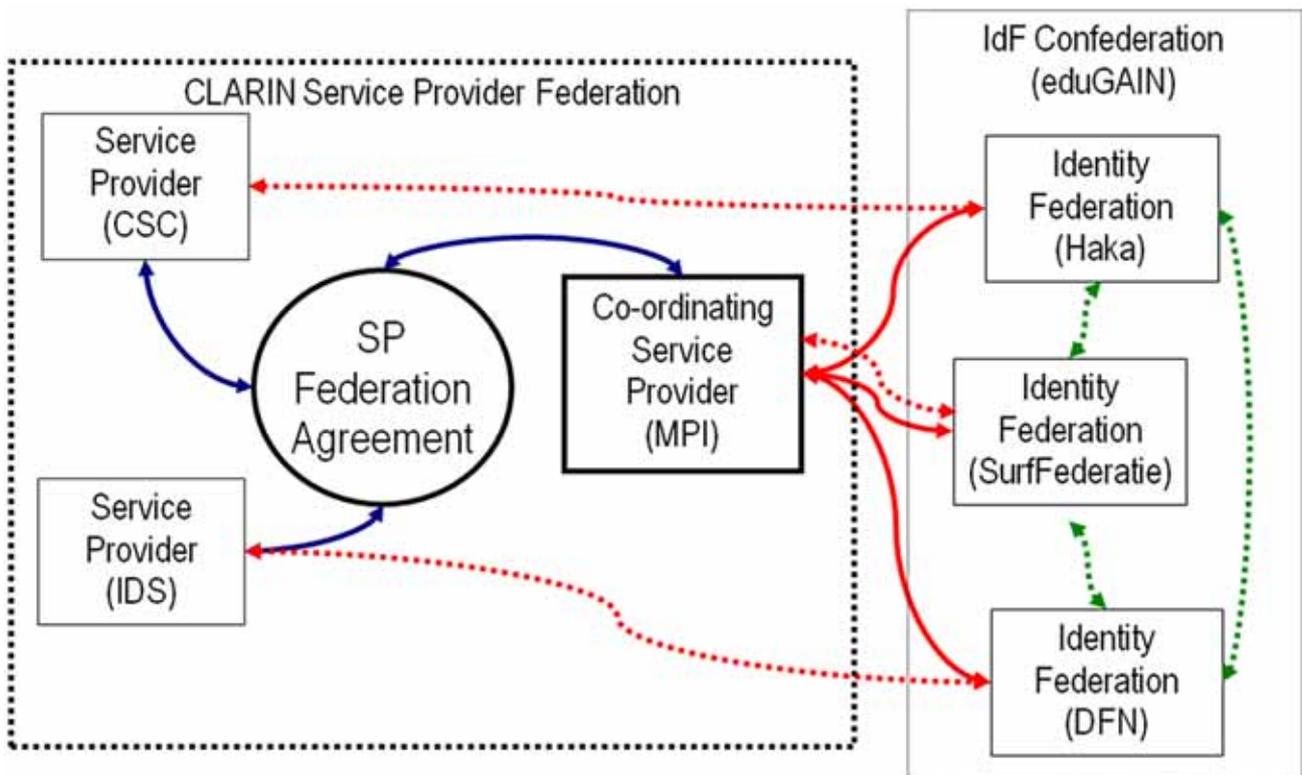
Blue lines: CLARIN Service Provider Agreements. Red lines: Service Provider Federation agreements with the Identity Federations.
Red dotted lines: Flow of Identity attributes

property rights of third parties, privacy issues, responsibilities towards third parties and the ethical dimension defined by Code of Conducts etc. Each service provider is entirely responsible for these matters on their own behalf. This is important to know for all depositors that the legal ground does not change.

Important characteristics of the Federation Agreements are a renewable two-year term, an intention to gain experience about various services, a centralized LRT resource registry, metadata harvesting, uniform PID services, compilation of virtual collections, common licensing models, a centralized EULA registry, etc. Recommendations on all of these will help service providers standardize their solutions. Each Service Provider is committed to comply with the obligations imposed by the currently valid national and EU legislation on personal data protection.

Why is this relevant for researchers?

Let us briefly state again why this federation will be relevant for researchers. A trust federation means that all parties involved rely on the proper authentication process at the researchers' home institute. This implies that the researcher has one identity and this one identity is for him/her the entry ticket to all services. When a researcher now wants to access resources from different repositories he/she only needs to authenticate once! **c**

Gabriela Haja
"A. Philippide" Institute of Romanian Philology, Romanian Academy
Eugen Munteanu
Centre for Biblical-Philological Studies, "Alexandru Ioan Cuza" University of Iaşi

Monumenta linguae Dacoromanorum (MLD) is a project that, starting 1986, lies at the heart of the partnership between Albert-Ludwigs-Universität of Freiburg, Germany and "Alexandru Ioan Cuza" University of Iaşi, Romania. The project will conduct research into the fundamental books of the Romanian ancient culture that were published in critical editions, using philological and IT tools. The series of these books opened in 1988 with the republishing of the first of the 22 projected volumes of the first complete Romanian translation of the *Bible* (1688). Up to the present moment, 8 volumes have been published. The initiators of the project were prof. dr. Paul Miron and prof. dr. Elsa Lueder from Albert-Ludwigs-Universität. The project has so far been carried out by academics of the Iaşi University, by philology specialists from the "A. Philippide" Institute of Romanian Philology – the Iaşi Branch of the Romanian Academy, together with researchers from Bucharest and Cluj-Napoca. In 2009, at prof. dr. Eugen Munteanu's initiative and with prof. dr. Elsa Lueder's consent, under the sponsorship of "Alexandru Ioan Cuza" University of Iaşi, the Centre for Biblical-Philological Studies *Monumenta Linguae Dacoromanorum* was founded, with the aim of perpetuating the traditional partnership between the universities of Iaşi and Freiburg. The critical edition of the 1688 *Bible* was completed, and a philological database with Romanian resources was created and made available online. Other projects include the editing, in paper and electronic formats, of a series of books that are fundamental for the history of Romanian written culture, and the publication of a number of studies to fill in the gaps from the Romanian historiography and philology or suggest new solutions for problems that, so far, have remained unsolved. Moreover, within the Centre for Biblical-Philological Studies, young researchers are instructed under the coordination of Professor Eugen Munteanu. The process of completion of the edition that began in 1988 has gathered six permanent employees and eighteen external collaborators from Bucharest, Craiova, Sibiu and Oradea.

The first steps in using IT tools in the research of ancient Romanian texts were taken at the University of Freiburg, under

International Networks of Centres in the Humanities

Paul Trilsbeek
J. Ringersma

The CLARIN way of viewing data and service centres as the backbone of future infrastructures for preservation and access is not the only attempt in this direction. Currently, we see at least two major international activities that have a close relationship with language resources. The DELAMAN (Digital Endangered LAnguages and Musics Archives Network) covers currently more than 20 centers worldwide that see it as their task to harmonize all efforts for the long-term preservation of all material about languages and cultures that will soon becoming extinct. It is a very important task to preserve this aspect of our cultural heritage and to give access to this material to future generations. All these centers see it as their task to digitize and archive the many recordings which are themselves highly endangered due to the deterioration of the physical substrates of the carrier (according to the UNESCO this seems to be true for 80% of the recordings on cultures and languages with ethnological background). DELAMAN is thus very important to synchronize standards, such as how to best encode video streams to guarantee quality and long-term accessibility, ways of archiving (safe replication of data) and ways to find funding for the digitization and curation of all the material that is still out there.

Another major initiative is centerNet which wants to cover relevant centers in the humanities worldwide. The initiative for this worldwide network comes from some of our colleagues in the US and it will work on similar dimensions such as helping each other to find opportunities and collaborators, to share tools and resources and to strengthen the position of centers within their own institutions as well as advancing digital humanities generally. A formal organizational form is currently being established where regional steering boards (one for Europe for example) take the role to organize the centers, push forward the ideas and build working groups on various issues. This is a fairly recent initiative, but it could have a large impact. Also in this area it is obvious that strong centers will be the backbone for the infrastructure and for advancements in e-Humanities from which ultimately all humanities researchers will profit. **c**

References

http://www.delaman.org/
http://digitalhumanities.org/centernet/

Monumenta Linguae Dacoromanorum.
The 1688 Bible Project

the coordination of Professor Elsa Lueder. Important progress was made, this time with the support of the researchers from the Faculty of Informatics of the University of Iaşi, through the completion of the volume that contains the first two *Kings' Books* (*Regum I, Regum II*, "Alexandru Ioan Cuza" University publishing house, 2008), under the coordination of Professor Paul Miron (Freiburg), of Professor Alexandru Andriescu (Iaşi) and of the researcher Gabriela Haja (Iaşi). The volume, the first to be published in electronic format, comprises, just like the ones mentioned above, next to the kirillic[1] texts of *I Kings* and *II Kings* of the 1688 *Bible* that is accompanied by the interpretive transcription in the Latin alphabet, also two new manuscripts from the 17th century, one belonging to Nicolae Milescu Spătarul and the other to Daniil Andrean Panoneanul (Fig. 1). The texts, aligned at the verse level, are accompanied by an up-to-date version of the biblical text, of vast philological commentaries, as well as by an exhaustive index of words and forms, automatically generated, subsequently validated and corrected by linguists (Fig. 2).

In parallel with the completion of the integral version of the 1688 *Bible* in the form suggested by MLD, the Centre for Biblical-Philological Studies will also carry out a more ample project, through which to create an electronic database for the research into the *Bible* in Romanian: *Tota Biblia Dacoromanorum (ToBiD)*. The latter supposes, on the one hand, a collection of all the direct sources – Romanian translations of the *Bible*,

Fig. 1. The Bible (1688) and the Interpretive Transcription

from the first texts (the 16th century) up to the current ones, and of all the indirect sources (monographs, studies, histories etc.). On the other hand, it also concerns the alignment of the direct sources at the verse level. Future projects will subsequently aim at aligning texts at the word level, according to the source of the translations -the Romanian translations traditionally took for a departure point Greek, Slavic, Latin originals; translations into modern languages are also to be added to the list – and to the stages of the historical evolution of the literary language, with vast interdisciplinary research possibilities.

Several Romanian *Bibles* are currently available on the Internet. They are, nevertheless, the result of individual initiatives that are

isolated, unprofessional and uncorrelated to one another.[2]

European funds will be used, starting 2010-2011, in order to implement the project. Contacts have already been established among the institutions wishing to carry out the project, such as: the Centre for Biblical-Philological Studies *Monumenta Linguae Dacoromanorum*, "Alexandru Ioan Cuza" University of Iaşi, the Romance Language Department – Albert-Ludwigs-Universität of Freiburg, the Faculty of Informatics – "Alexandru Ioan Cuza" University of Iaşi, the "A. Philippide" Institute of Romanian Philology – the Iaşi Branch of the Romanian Academy, the list remaining open for persons that are interested to colaborate in the field of computer-assisted philological study of the biblical texts in Romanian. **C**

Fig. 2. Title Page of the Electronic Format of the Last Volume and Index with Links to the Three Translation Versions from the 17th Century (screenshots)

Freedom of use of copyrighted works for academic purposes
NEERI Manifesto

Helsinki, October 2009

The emergence of e-Science

The emergence of new research methods and practices that exploit advanced computational resources, data collection, and scientific instruments, made easily accessible within a distributed information network that combines all such resources into a virtually singular entity – that is e-Science – will revolutionize how science is done. In this era of e-Science, researchers work in a global, borderless world where data can be accessed, combined and processed across scientific disciplines and geographical barriers in a multitude of ways. In order to make this new way of creating scientific knowledge possible, there is one big barrier, which has to be torn down – the current inflexible copyright regime.

The importance and value of data in research

Knowledge can be regarded as a global public good which is non-exclusive and non-rival. The use of knowledge by any one person does not prevent others from

using it, nor does the use of knowledge diminish it – quite the contrary, using existing knowledge cumulatively creates new knowledge. Treating knowledge as private intellectual property turns knowledge exclusive and thus retracts it

CLARIN Team wins a Special Award 'Intel Education'

Marius Zbancioc
Member of SRoL Group

SRoL is a project affiliated to CLARIN, which is building an annotated database for the spoken Romanian language. The SRoL Web-based spoken language repository and tool collection was developed over several years by the collaboration of groups from the Institute for Computer Science of the Romanian Academy, CERFS Excellence Center in "Gheorghe Asachi" Technical University of Iasi and staff of the discipline of Language Technology, Computer Science Faculty, "Al.I. Cuza" University. The web site includes about two thousand voice recordings grouped into sections like "Basic sounds of the Romanian language", "Emotional voices", "Specific language processes", "Pathological voices", "Comparison of natural and synthetic speech", and "Gnathophonics and Gnathosonics". The recordings are annotated and documented according to proprietary methodology and protocols. Moreover, the site includes extended documentation on the Romanian language, and on speech technology, moreover tools produced by us, for voice analysis. The resources are a part of the CLARIN European Network for Language Resources. The resources and tools are useful for (virtual) learning of the phonetics of the Romanian language, and of medical subjects related to voice.

At the 4th International Conference on Virtual Learning – ICVL 2009 (which was held under the auspices of EYCI- the European Year of Creativity and Innovation 2009, The European INTUITION Consortium, the Romanian Ministry of education and Research, the National Authority for Scientific Research), the SRoL team pre-

from the public sphere. In particular, research data produced by public funding should be regarded as a public good, openly available to the maximum extent possible at the global level. The production, sharing and open dissemination of primary research data are essential attributes of, and inputs to, modern systems of scientific research and innovation.

Current state of regulating use of copyrighted works for academic purposes

The European Commission has in its *Green Paper on Copyright in the Knowledge Economy* recognized that current Copyright law within the EU is not harmonized nor does it support creating pan-European research networks.

The EU Directive *Copyright in the Information Society* specifies an exhaustive but unfortunately nonmandatory list of exceptions, which the EU Member States may opt to implement. Importantly, these exceptions include the option of allowing the "[use of copyrighted works] for the sole purpose of illustration for teaching or scientific research", as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the noncommercial purpose to be achieved.

However, the European Commission acknowledges that this exception has been implemented in diverse ways among the Member States and typically in a narrow sense. This goes directly against the free movement of knowledge within the Common Market which is emphasized in the Lissabon goals of the European Union. Indeed, since the free circulation of knowledge is regarded as one of the cornerstones for the future development of the knowledge economy in Europe, it is essential that the public interest with respect to the creation and availability of scientific knowledge is governed by similar rules throughout the community.

What we need is a detailed but flexible copyright exception that provides the Member States the clear path to adapt their legal framework to the changes in technology, in the practices of producing scientific knowledge, and in the markets, and which takes into account the public interest.

Common ESFRI proposal

The creation as well as the efficient functioning of any pan-European research infrastructure is virtually impossible if the exemptions for academic use are not harmonized within EU, and eventually world-wide.
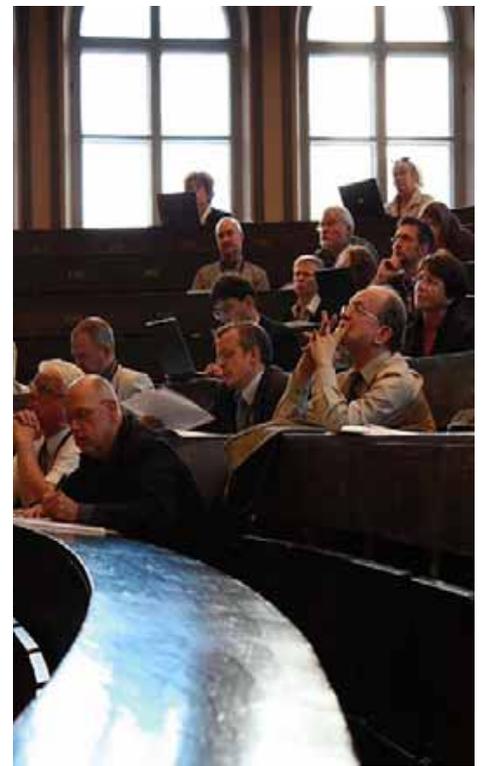
Therefore, we urge the legislatures at the European level and at the national level to harmonize copyright law in such a way that the free use of copyrighted works for academic purposes, which does not unreasonably prejudice the legitimate interests of the right-holders, is permitted in all member states. C

---

sented the project SRoL and several applications in language learning and voice technology classes.

During the closing session of the conference, two awards were offered:

- The Excellence Award "Intel Education", accorded to an Italian team for the paper: "E-portfolio and semantic web to support informal learning in social network environment"
- The Special Award "Intel Education" presented for the paper "SRoL – Web-based Resources and Tools used for Language and Language Technology e-Learning", by Monica Silvia Feraru, Horia-Nicolai Teodorescu, from the Institute for Computer Science, Romanian Academy, Iasi Branch. The corresponding paper was publised in the volume of the conference Virtual Learning – Virtual Reality, Proc. 4th International Conference on Virtual Learning, ICVL 2009, (Bucharest University Press, ISSN: 1844-8933, pp. 119-127).

This is not the first public success of the SRoL. The SRoL resources have been recognized by several bodies, beyond the scientific publications that included papers on SRoL. ORDA (the Romanian Office for Authorship Rights) registered the original recordings. The SRoL received a gold medal and media attention at the INVENTICA 2009 fair for inventions and creativity. Also, the website of Ambassade de France in Romania briefly described in its Bulletin the SRoL site and its use in education (http://www.bulletins-electroniques. com/actualites/58811.htm).

The SRoL corpus is currently used in teaching and laboratory activities in the class "Speech Technology" given for the master degree in "Computational Linguistics" at the Faculty of Computer Science, "Al.I. Cuza" University of Iaşi. At the international EUROLAN 2007 summer school, the SRoL site was used to present "Traces of emotion, intentions and meaning in spoken Romanian" (http://eurolan.info. uaic.ro/html/profs/HNTeodorescu.html)

The Technical University "Gheorghe Asachi" of Iasi intends to use SRoL in helping foreign students enrolled at this university learn Romanian pronunciation. C

Ecoystem of Infrastructures

Peter Wittenburg
MPI, Nijmegen
Daan Broeder
MPI, Nijmegen
Marc Kemps-Snijders
MPI, Nijmegen

This new term "Eco-System of Infrastructures" is certainly a new buzzword which is currently en vogue. We had to find out ourselves what it means and why it is becoming so important. It certainly is a fuzzy term, invoking visions of a dynamic environment and mutual dependencies, but it is clearly a sign of the current "gold-rush" atmosphere we are all participating in – at least those who are enthusiastic of building research infrastructures such as CLARIN, DARIAH, LIFEWATCH etc and those who are building so-called generic e-Infrastructures in Europe. While the e-Infrastructures such as GEANT (network), EGEE/ EGI (grid), DEISA/ PRACE (high performance computing) have partly already been started some years ago, the research infrastructures are all babies of the ESFRI process.

Why was ESFRI step ahead

The ESFRI roadmap result was an enormous step ahead for many disciplines. Why is this so?

1) We could start a concerted action in many disciplines all over Europe – something which would have been almost impossible otherwise, since the national interests and discussion processes are much too different.

2) This concerted action has already now changed so many minds in the various disciplines in the direction of changing the culture in which research is being carried out. Researchers are much more aware of the fact that exchange, interoperability, cross-fertilization etc. will become important for remaining at the leading edge of research in areas which are dependent on data and advanced technology. Almost everyone in the research domain has now heard about ESFRI and has been exposed to some of the discussions taking place.

3) Finally, we received concrete funding in this concerted action to make steps ahead, to interact with many of our colleagues, to develop skills in particular of young people and to get a much better understanding of the obstacles that might hamper forwardleaps towards an e-Research scenario.

So many of us started with great enthusiasm, of course, underestimating the size of the task ahead of us and solving all similar problems, at least to a certain extent. On the other hand e-Infrastructures certainly suffer from the fact that, allthough they have received years of funding, and can benefit from established basic services such as high-speed networks, they also led to continuous invention of new paradigms and ideas which are not necessarily required by the research communities.

The essential four questions

What we currently can see is that experts are recognizing these phenomena. The essential four questions are:

1) How can we make sure that the e-Infrastructures will offer exactly those services that are required by the communities?

2) How can we make sure that problems that are shared by a few or even all of the research infrastructures will be solved in a generic way?

3) Have we missed essential services that have not been addressed so far?

4) How can we make sure that at every layer we arrive at lean services that can be financed and maintained for a longer period?

The optimal answer to these questions will certainly describe a successful ecosystem of infrastructures, and we immediately understand that it will be a long path to come to a satisfying equilibrium. There are many reasons why this transition needs to take some time; here we just want to discuss one aspect which is often ignored in discussions. For researchers "trust" is the keyword for adopting new methods. Without sufficient support from domain-oriented experts who can translate new technologies into the cultures and traditions of the disciplines it will be hard to build trust. Training sufficient experts of this sort can only be done by carrying out concrete projects (learning by doing) and only these experts will be able to communicate with the e-Infrastructure experts about what is really needed. All this will take time, no matter what procedures are chosen.

Intensify our collaborations and interactions

We have already argued that in this light the result of the ESFRI roadmap process must be seen as an enormous step for advancing the research disciplines in Europe. Currently we are all finishing the preparation phase and also the e-Infrastructures are undergoing a critical evaluation. It is obvious that we need to intensify our collaborations and interactions. The NEERI conference and also the EGI user forum were already good starting points. Currently, there are also discussions at the level of the European Commission about how to stimulate these interactions and of how to fill existing gaps. It is obvious that in the domain of e-Infrastructures we miss a systematic European approach to tackling long-term preservation and accessibility of research data. It seems that the EC is willing to close this gap by substantial funds. Obviously this can only be done by bringing together some "structured" communities, i.e. communities that have clear requirements, and some strong data centers with deep IT knowhow. Of course it would not be feasible to create a project covering many of these data centers across all research infrastructures. Therefore, one can speak about an excellent complementary step from the EC to define clusters of thematically close research infrastructures and launch calls where these clusters can actively study commonalities and find joint solutions in the area of data related services. Also these clusters should involve experts from the e-Infrastructures to include the experience which is already out there.

We expect that this dualistic approach will help us to come closer to a balanced ecosystem of infrastructures. However, we will need time to define the various roles and layers of lean services. **C**

Recruiting and training tomorrow's researchers

CLARA project started

Koenraad de Smedt
University of Bergen

The construction and exploitation of a research infrastructure for language is dependent on the availability of researchers with specialized competencies and skills. To prepare for this need, a Marie Curie Initial Training Network has been initiated under the name CLARA (Common Language Resources and their Applications). CLARA is recruiting 17 researchers for initial training at the following institutions:

1. University of Bergen
2. Eberhard-Karls University (Tübingen)
3. Tilde (Riga)
4. Copenhagen University
5. University Pompeu Fabra (Barcelona)
6. Helsinki University
7. Charles University (Prague)
8. Norwegian School of Economics and Business Administration (Bergen)
9. Max Planck Institute for Psycholinguistics (Nijmegen)

The training program, consisting of supervision, courses and summer schools, also involves the following associate partners: Hungarian Academy of Sciences, Comperio, TEMIS Germany, ELRA, University of Zagreb, Uni Digital (formerly Unifob AKSIS) and Lexical Computing.

Vacancies for researchers are in various research tasks and for durations of 24 to 36 months.

All positions require mobility, that is, are open only to researchers coming from different countries than the country where they will be appointed. There are two types of positions:

1. Early Stage Researcher (ESR): postgraduate without PhD degree and with less than 4 years of research experience (after obtaining master diploma or equivalent)
2. Experienced Researcher (ER): researcher with less than 5 years of research experience and with PhD or at least 4 years of research experience

Vacancies are announced on the CLARA website (http://clara.uib.no) and on Euraxess, together with instructions on how to apply. C

Social Sciences and Humanities Futures Workshop

Brussels

October 22-23, 2009

Marko Tadić
Editor

The SSH Futures workshop that concluded the FP6 project "The Future of Social Sciences and Humanities" was held in Brussels from 22 to 23 October 2009. The project CLARIN and it role in building research infrastructures for SSH was presented by Marko Tadić. Although the stress of the workshop was primarily on Social Sciences and much less on Humanities, since the participants were predominantly from sociology, social policy, political sciences, social anthropology, etc. So this setting was both appropriate and inappropriate for our presentation. Inappropriate because RI for SSH based mostly on LRT was in a way out of the primary scope of the workshop, but also appropriate because just because of this contrast the impression or impact was even stronger. Within the conference there was an interesting section IVc "Complex methodologies" where some computational methods in SSH were mentioned and in the discussion they were made aware of CLARIN RI and LRT community proper. C

Digital Resources and Knowledge Sharing

INFuture2009 conference Zagreb

November 4-6, 2009

Sanja Seljan
University of Zagreb, Faculty of Humanities and Social Sciences

The second conference *The Future of Information Sciences – INFuture 2009: Digital Resources and Knowledge Sharing* was held from 4 to 6 November in Zagreb. The conference was co-organised by the Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb and IBM Croatia.

The biannual conference *The Future of Information Sciences – INFuture* aims to bring together researchers, professionals, businessmen and project managers from the broad field of information sciences and related professions. The objective of the conference is to provide a platform for discussing theoretical and practical issues in the field of information sciences.

Language Technologies in the Humanities

At the conference about 140 participants form different sectors and countries participated: form Italy, Germany, the Netherlands, Belgium, United Kingdom, Slovenia, Bosnia and Herzegovina, Serbia, India, Iran, United States of America, China and Croatia.

About 80 papers were presented in different sessions elaborating through multidisciplinary approaches on the needs, problems and possible models for digitization and heritage preservation, with special focus on language technologies and their application in the Humanities (e.g. digitization and preservation of cultural heritage, language identity, language processing for Croatian, use of open-source solutions, etc.).

The conference was divided into following sections: (1) Invited papers, (2) Digitization and preservation, (3) Using of open-source solutions in cultural heritage, (4) Language technologies, (5) Using information resources in research, education and presentation, (6) Virtual environments in education, (7) E-services, e-government and business applications, and (8) Knowledge management.

CLARIN's contribution to national RI roadmap

Steven Krauwer, as keynote speaker, presented the CLARIN project (the Faculty of Humanities and Social Sciences, University of Zagreb is a consortium partner of the CLARIN project); Alexander Fraser gave a report on bitext parsing; Chiara Cirinna and Maurizio Lunghi demonstrated the importance of long term preservation of digitized materials; Bob Bater and David Bawden considered the sharing of knowledge and vision for the 21st century.

A *round table discussion* on two FP7 projects – DARIAH and CLARIN, was held with about 50 participants (6th November). The title of the round table was *Development of a national strategy for digital preservation* and the foundations of national roadmap for RI were laid down.

Two pre-event workshops were held on 2nd November: *Knowledge: Making it Work* and *The use of Web 2.0 Tools for e-Learning* attended by about 50 participants.

The participants hoped that scientific and cultural cooperation at the international level would be enabled through digitization, heritage preservation in the European context, through preservation of the Croatian national identity, use of open-source solutions in research and education, and through language technologies and multimedia content. **c**

Tamás Váradi
CLARIN EB member

As one of the few non Indo-European languages on the European landscape, the Hungarian language may be a linguistic island, but Hungarian language technology is certainly well embedded in the European language technology scene.

Members of HUN-CLARIN

The Hungarian CLARIN Network (HUN-CLARIN) was formed by a number of centres which had been successfully collaborating with each other in a string of national LT projects. They have complementary expertise and resources: the *Reseach Institute for Linguistics, Hungarian Academy of Sciences* (HASRIL), is the developer of basic resources, most notably the Hungarian National Corpus, and acts as the coordinator; *Szeged University* is strong in machine learning techniques and has developed the first Hungarian treebank, the Szeged Corpus; the *Media Research Centre of the Technical University* has dveloped a series of open source tools and resources including the widely used HunAlign and the popular Hunglish Hungarian-English parallel corpus; *MorphoLogic Ltd.* represents a commerical company that engages in HLT research and resources development as well, their main area of activity

Magyar Tudományos Akadémia
Nyelvtudományi Int

This is a **beta test** of our sentence search. The sentences are t
sentences have been aligned automatically. We use the hunmorp
found even in inflected form.

please post your comments and questions to the open mailing lis

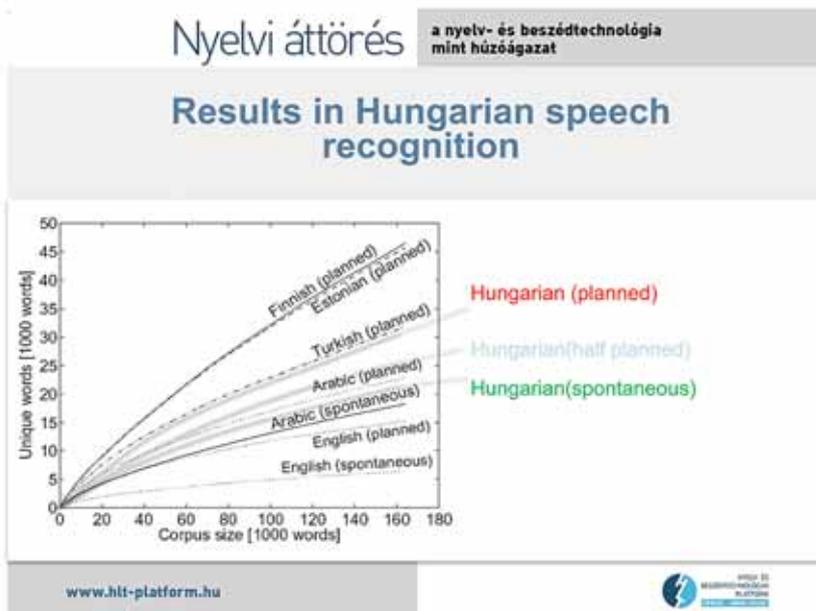Download firefox plug-in

Hungarian: [   ]   English: under pressure

Results 1 - 20 of 42069 from the Hunglish corpus

Legal texts	- a negatív légnyomás alatt tartást,
Literature	Gaal erősen összpontosított.
Movie subtitles	Nagy nyomás alatt volt.
Movie subtitles	Nem helyezek rád nagy nyomást, iga
Movie subtitles	Stresszes élete lehet.
Legal texts	- Gázok: sűrített, cseppfolyósított, va alatt feloldott,
Literature	Essie egy túlnyomásos bura alatt lehe
Literature	- Caramel a kalapját gyűrögette.

Progress report on the Hungarian CLARIN Network

being proofing tools and machine translation. The HUN-CLARIN network can boast with Hungary's leading speech technology centre, the *Department of Telecommunication and Media Informatics of the Technical University*, a huge department which has developed pioneering and marketable speech synthesis and recognition technology. The *Institute for Psychological Research* joined HUN-CLARIN as a result of their success in applying language technology in the content analysis of narrative psychological accounts, a collaboration reported in the first issue of the CLARIN Newsletter.

The HUN-CLARIN community enjoys the endorsement of the National Office for Research and Technology. This endorsement – intitally embodied in a Letter of Support submitted to the CLARIN consortium – has since been converted into material support through a Call that NORT has issued for the creation of national technology platforms late 2007. This was an excellent opportunity, members of the HUN-CLARIN community felt, to promote HLT in Hungary and submitted a proposal that was accepted despite the fact that HLT is very little known in the general public or even, for that matter, among the tech-

nology savvy crowd in Hungary. The *National Language and Speech Technology Platform*, joins together an equivalent number of small and medium businesses (SMEs) with the HUN-CLARIN members listed above. As a project, the Platform is supposed to work out their Strategic Research Agenda and a corresponding Implementation Plan. In addition, the grant enables the Platform to promote HLT among the policy makes, in the media and the general public. Although, the Platform is expected to be oriented more towards business applications its mission encompasses a lot of the objectives of CLARIN itself. The Platform welcomed Steven Krauwer to its opening conference, who gave a rousing opening plenary lecture. The Strategic Research Agenda was completed and inaugurated at the second conference last December.

Official policy for RI

One drawback in gaining financial support for developing research infrastructure was the lack of official policy for research infrastructures in Hungary. The ESFRI process has finally changed this. In 2009 large-scale work began to compile the Hungarian Research Infrastructure Roadmap, codenamed NEKI-

FUT. The review process follows the design of the original European initiative e.g. just like the ESFRI Roadmap, it is conducted in three working. One peculiar feature of the first round of the Roadmap survey that individual resources and tools (whatever their size) were invited to enter along with whole networks like HUN-CLARIN. The first phase of the review process is over and I am pleased to find that HUN-CLARIN has emerged as the strongest candidate in the SSH area. We are about to enter the second phase, prior to which the resource and tool providors (research infrastructure operators in the NEKI-FUT world) are formally requested to form a consortium and compete together.

What is at stake in NEKIFUT? Well, plans have not crystallized yet but pledges have been made that those who are accorded the title Strategic Research infrastructure, can expect to receive regular support to build and maintain the infrastructure.

At present, HUN-CLARIN is a strong contester – giving us strong hope (no more than that in the midst of severe fiscal and economic measures to counter with an impending general election in the spring) that Hungary will continue its present role within CLARIN. Watch this space! C

Join CLARIN

The CLARIN project is a combination of Collaborative Projects and Coordination and Support Actions, registered at the EU under the number FRA-2007-2.2.1.2. It started with the preparatory phase in 2008 that will make the grounds for the next phases and it will cover the generic, language independent activities. In order to do our work properly we have to rely on a much wider circle than just the formal consortium partners in the project. For this reason we have opened up all our project working groups for participation by organizations that are not part of the consortium.

Members

Austria (NCP: Gerhard Budin) University of Graz (Graz): Austrian German Research Centre (C:Rudolf Muhr); Institut für Romanistik (C:Stefan Schneider)

Austrian Academy of Sciences (Vienna): Austrian Academy Corpus (C:Christoph Benda); Department of Linguistics and Communication Research (C:Sabine Laaha); Institute of Lexicography of Austrian Dialects and Names (C:Eveline Wandl-Vogt)

Secure Business Austria (Vienna): (C:Edgar R. Weippl)

University of Vienna (Vienna): Center for Translation Studies (C:Gerhard Budin)

Belgium (NCP: Ineke Schuurman) University of Antwerp (Antwerp): Center for Dutch Language and Speech (C:Walter Daelemans)

Vrije Universiteit Brussel (Brussels): Laboratory for Digital Speech and Audio Processing, Department of Electronics and Information Processing (C:Werner Verhelst)

Gent University (Gent): Digital Speech and Signal Processing research group at the Electronics and Information Systems department (C:Jean-Pierre Martens)

University College Ghent (Gent): Faculty of Translation Studies, Language and Translation Technology Team (C:Veronique Hoste)

Katholieke Universiteit Leuven (Leuven): Center for Computational Linguistics (C:Frank Van Eynde); ESAT-PSI/Speech (C:Patrick Wambacq); Language Intelligence & Information Retrieval (C:Marie-Francine Moens)

Katholieke Universiteit Leuven (Leuven - Kortrijk): iTec (Interdisciplinary research on Technology, Education & Communication) (C:Hans Paulussen)

Bulgaria (NCP: Kiril Simov) University of Plovdiv (Plovdiv): Faculty of Mathematics and Informatics (C:Veska Noncheva)

Bulgarian Academy of Sciences (Sofia): Department of Computational Linguistics, Institute for Bulgarian Language (C:Svetla Koeva); Institute for Parallel Processing of Bulgarian Academy of Sciences (Sofia): Linguistic Modelling Department (C:Kiril Simov); Institute of Mathematics and Informatics, Bulgarian Academy of Sciences (Sofia): Mathematical Linguistics Departement (C:Ludmila Dimitrova)

St. Cyril and St. Methodius University (Veliko Turnovo): (C:Boryana Bratanova)

Croatia (NCP: Marko Tadić) Institute of Croatian Language and Linguistics (Zagreb): (C:Damir Cavar)

University of Zagreb (Zagreb): Department of Linguistics, Faculty of Humanities and Social Sciences (C:Marko Tadić); Zagreb University Computing Center (C:Zoran Bekić)

Cyprus (NCP: -) Cyprus College (Nicosia): Research Center (C:Antonis Theocharous)

Czech Republic (NCP: Eva Hajičová) Masaryk University (Brno): Faculty of Informatics (C:Aleš Horák)

Charles University (Prague): Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics (C:Eva Hajičová)

The Institute of the Czech Language, Czech Academy of Sciences (Prague): The Institute of the Czech Language (C:Karel Oliva)

Denmark (NCP: Bente Maegaard) Copenhagen Business School (Copenhagen): Department of International Language Studies and Computational Linguistics (C:Peter Juel Henrichsen)

Dansk Sprognævn - Danish Language Council (Copenhagen): (C:Sabine Kirchmeier-Andersen)

Society for Danish Language and Literature (Copenhagen): (C:Jørg Asmussen)

The National Museum of Denmark (Copenhagen): (C:Birgit Rønne)

The Royal Library (Copenhagen): (C:Anders Conrad)

University of Copenhagen (Copenhagen): Centre for Language Technology, Faculty of Humanities (C:Bente Maegaard)

University of Southern Denmark (Kolding): Faculty of Humanities (C:Johannes Wagner)

Estonia (NCP: Tiit Roosmaa) University of Tartu (Tartu): Institute of Computer Science (C:Tiit Roosmaa)

Finland (NCP: Kimmo Koskenniemi) CSC - the Finnish IT Center for Science (Espoo): (C:Pirjo-Leena Forsström)

Lingsoft Inc. (Helsinki): (C:Juhani Reiman)

The Research Institute for the Languages of Finland (Helsinki): (C:Toni Suutari)

University of Helsinki (Helsinki): Department of General Linguistics (C:Kimmo Koskenniemi)

University of Joensuu (Joensuu): Department of Foreign Languages and Translation Studies (C:Jussi Niemi)

University of Oulu (Oulu): Faculty of Humanities, Finnish Language (C:Marketta Harju-Autti)

University of Tampere (Tampere): Faculty of Information Sciences, Department of Information Studies and Interactive Media (C:Eero Sormunen)

France (NCP: Jean-Marie Pierrel) Centre de ressources pour la documentation de l'oral (Aix-en-Provence) (C:Bernard Bel)

National Center for Scientific Research (CNRS) (Marseille): Laboratoire d'Informatique Fondamentale de Marseille (LIF-CNRS) (C:Michael Zock)

Centre National de Ressources Textuelles et Lexicales (CNTRL) (Nancy): (C:Bertrand Gaiffe)

National Center for Scientific Research (CNRS) (Nancy): Analyse et Traitement Informatique de la Langue Française (ATILF) (C:Jean-Marie Pierrel)

National Center for Scientific Research (CNRS) (Orsay): Institute for Multilingual and Multimedia Information (IMMI-CNRS) (C:Joseph Mariani)

Evaluations and Language resources Distribution Agency (ELDA) (Paris): (C:Khalid Choukri)

National Center for Scientific Research (CNRS) (Paris): Traitement Électronique des Manuscrits et des Archives (TELMA/DIS) (C:Florence Clavaud)

Université Paris 4 Sorbonne (Paris): Centre de linguistique théorique et appliquée (CELTA) (C:Andre Wlodarczyk)

Université de Strasbourg (Strasbourg): Equipe de recherche LiLPa (Linguistique, Langues, Parole) (C:Amalia Todirascu)

National Center for Scientific Research (CNRS) (Vandoeuvre les Nancy): L'Institut de l'Information Scientifique et Technique (INIST-CNRS) (C:Fabrice Lecocq)

University Paris Est/Paris 12 (Vitry Sur Seine): LISSI Laboratory (C:Yacine Amirat)

Germany (NCP: Erhard Hinrichs) University of Augsburg (Augsburg): Philologisch-Historische Fakultät (C:Ulrike Gut)

Berlin-Brandenburg Academy of Sciences (Berlin): (C:Alexander Geyken)

Humboldt-University Berlin (Berlin): Institut für deutsche Sprache und Linguistik (C:Anke Lüdeling)

Technische Universität Darmstadt (Darmstadt): Ubiquitous Knowledge Processing (UKP) Lab (C:Iryna Gurevych)

TU Dortmund University (Dortmund): Institute for German Language and Literature (C:Michael Beißwenger)

Universität Duisburg-Essen (Essen): Fakultät Geisteswissenschaften / Germanistik / Linguistik (C:Bernhard Schröder)

University of Frankfurt/Main (Frankfurt/Main): Comparative Linguistics Department (C:Jost Gippert)

University of Giessen (Giessen): Institut für Germanistik (C:Henning Lobin)

DFG-Project "Language Variation in Northern Germany" (Sprachvariation in Norddeutschland - SiN) (Hamburg): (C:Ingrid Schröder)

University of Hamburg (Hamburg): Faculty for Language Literature and Media, Arbeitsstelle "Computerphilologie" (C:Cristina Vertan); Fakultät für Geisteswissenschaften, Fachbereich Sprache, Literatur, Medien (C:Angelika Redder); Institute of German Sign Language and Communication of the Deaf (C:Thomas Hanke); SFB 538 Multilingualism (C:Thomas Schmidt)

University of Heidelberg (Heidelberg): Computational Linguistics Department (C:Anette Frank)

University of Cologne (Köln): Institut für Linguistik - Phonetik (C:Dagmar Jung)

Max Planck Institute for Evolutionary Anthropology (Leipzig): Department of Linguistics (C:Hans-Jörg Bibiko)

University of Leipzig (Leipzig): Institut für Informatik, Abteilung Automatische Sprachverarbeitung (C:Codrina Lauth)

Institut für Deutsche Sprache (Mannheim): (C:Marc Kupietz)

Westfälische Wilhelms-Universität Münster (Münster): Institut für Allgemeine Sprachwissenschaft (C:Gabriele Müller)

University of Potsdam (Potsdam): Department of Linguistics (C:Manfred Stede)

German Research Center for Artificial Intelligence (Saarbrücken): Language Technology Lab (C:Thierry Declerck)

University of Stuttgart (Stuttgart): Institut für Maschinelle Sprachverarbeitung (C:Ulrich Heid)

Universität Trier (Trier): Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften (C:Andrea Rapp)

Universität Tübingen (Tübingen): Asien-Orient-Institut (C:Ulrich Apel); Seminar für Sprachwissenschaft (C:Erhard Hinrichs)

Greece (NCP: Stelios Piperidis) Institute for Language and Speech Processing (Athens): Department of Language Technology Applications (C:Stelios Piperidis)

Hungary (NCP: Tamás Váradi) Hungarian Academy of Sciences (Budapest): Research Institute for Linguistics (C:Tamás Váradi); Institute for Psychological Research of the Hungarian Academy of Sciences (Budapest): (C:Bea Ehmann)

Budapest University of Technology and Economics (Budapest): Department of Sociology and Communications, Media Research Center (C:Peter Halacsy)

Department of Telecommunication and Media Informatics, Laboratory of Speech Acoustics (C:Klára Vicsi)

MorphoLogic Ltd. (Budapest): MorphoLogic Ltd. (C:László Tihanyi)

University of Szeged (Szeged): Department of Informatics, Human Language Technology Group (C:Dóra Csendes)

Iceland (NCP: Eiríkur Rögnvaldsson) Icelandic Centre for Language Technology (Reykjavik): (C:Eirikur Rögnvaldsson)

University of Iceland (Reykjavik): Institute of Linguistics (C:Eirikur Rögnvaldsson)

Ireland (NCP: -) National University of Ireland (Galway): Department of English (C:Sean Ryder)

Israel (NCP: -) Technion-Israel Institute of Technology (Haifa): Computer Science Department (C:Alon Itai)

Italy (NCP: Nicoletta Calzolari) European Academy Bozen/Bolzano (Bolzano): Institute for Specialised Communication and Multilingualism (C:Andrea Abel)

Università di Pavia (Pavia): Dipartimento di Linguistica Teorica e Applicata (C:Andrea Sansò)

National Research Council (Pisa): Istituto di Linguistica Computazionale (C:Nicoletta Calzolari)

University of Rome "Tor Vergata" (Rome): Department of Computer Science (C:Fabio Massimo Zanzotto)

Latvia (NCP: Inguna Skadina) Tilde Language Technologies (Riga): Tilde Language Technologies (C:Andrejs Vasiljevs)

University of Latvia (Riga): Institute of Mathematics and Computer Science (C:Inguna Skadina)

Lithuania (NCP: Ruta Marcinkieviciene) Vytautas Magnus University (Kaunas): Center of Computational Linguistics (C:Ruta Marcinkeviciene)

Institute of the Lithuanian Language (Vilnius): (C:Daiva Vaisniene)

Luxembourg (NCP: -) European Language Resources Association (ELRA) (Luxembourg): (C:S.Piperidis/K.Choukri)

Malta (NCP: Mike Rosner) University of Malta (Malta): Department of Computer Science (C:Michael Rosner)

Netherlands (NCP: Jan Odijk) Meertens Institute (Amsterdam): Meertens Institute (C:H.J. Bennis)

Univerity of Amsterdam (Amsterdam): Intelligent Systems Lab Amsterdam (ISLA) (C:Maarten de Rijke)

Vrije Universiteit Amsterdam (Amsterdam): Computational Lexicology, Faculteit der Letteren (C:Piek Vossen)

Data Archiving and Networked Services (Den Haag): (C:Henk Harmsen)

Huygens Instituut KNAW (Den Haag): (C:K.van Dalen-Oskam)

University of Twente (Enschede): Human Media Interaction Group, Department of Electrical Engineering, Mathematics and Computer Science (C:Roeland Ordelman)

University of Gröningen (Gröningen): Faculty of Arts, Center for Language and Cognition (C:Wyke van der Meer)

Digital Library for Dutch Literature (Leiden): (C.C.A. Klapwijk)

Institute for Dutch Lexicology (Leiden): Instituut voor Nederlandse Lexicologie (C:Remco van Veenendaal)

Universiteit Leiden (Leiden): Leiden University Centre for Linguistics, Faculty of Humanities (C:Jeroen van de Weijer)

Max Planck Institute for Psycholinguistics (Nijmegen): (C:Peter Wittenburg)

Radboud University (Nijmegen): Centre for Language and Speech Technology (C:L. Boves / N. Oostdijk); Centre for Language Studies (C:Pieter Muysken)

Tilburg University (Tilburg): ILK Research Group, Department of Communication and Information Sciences, Faculty of Humanities (C:Antal van den Bosch)

University of Utrecht (Utrecht): Utrecht Institute of Linguistics OTS, Faculty of Humanities (C:Steven Krauwer)

Norway (NCP: Koenraad De Smedt) Norwegian School of Economics and Business Administration (NHH) (Bergen): (C:Gisle Andersen)

Unifob AS (Bergen): (C:Eli Hagen)

University of Bergen (Bergen): Language Models and Resources group (C:Koenraad de Smedt)

SINTEF (Oslo): (C:Diana Santos)

The Language Council of Norway (Oslo): (C:Torbjørg Breivik)

The National Library of Norway (Oslo): (C:Kristin Bakken)

University of Oslo (Oslo): Department of Linguistics and Nordic Studies, Faculty of Humanities (C:Janne Bondi Johannessen)

University of Tromsø (Tromsø): Det humanistiske fakultet (C:Trond Trosterud)

Norwegian University of Science and Technology (Trondheim): Department of Electronics and Telecommunications (C:Torbjørn Svendsen)

Poland (NCP: Maciej Piasecki) University of Lodz (Lodz): Institute of English Language (C:Piotr Pezik)

Polish Academy of Sciences (Warsaw): Institute of Computer Science, Department of Artificial Intelligence (C:Adam Przepiórkowski); Institute of Slavic Studies (C:Violetta Koseska-Toszewa)

Polish-Japanese Institute of Information Technology (Warsaw): (C:Krzysztof Marasek)

University of Wroclaw (Wroclaw): Instytut Informatyki Stosowanej (C:Maciej Piasecki)

Wroclaw University of Technology (Wroclaw): Institute of Applied Informatics (C:Maciej Piasecki)

Portugal (NCP: Antonio Branco) Universidade Católica Portuguesa (Braga): Centro de Estudos Filosóficos e Humanísticos (C:Augusto Soares da Silva)

University of Minho (Braga): Centro de Estudos Humanísticos (C:Pilar Barbosa)

New University of Lisbon (Caparica): Faculdade de Ciências e Tecnologia (C:José Gabriel Pereira Lopes)

Instituto de Telecomunicações (Coimbra): Pólo de Coimbra (C:Fernando Perdigão)

University of Coimbra (Coimbra): Centro de Esudos de Linguística Geral e Aplicada (CELGA) (C:Cristina Martins); Centro de Investigação do Núcleo de Estudos (C:José Augusto Simões Gonçalves Leitão)

Universidade de Évora (Évora): School of Sciences and Technology (C:Paulo Quaresma)

INESC-ID, Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento em Lisboa (Lisboa) (C:Nuno Mamede)

Instituto de Linguística Teórica e Computacional (Lisbon): (C:Margarita Correia)

New University of Lisbon (Lisbon): Centro de Linguística (C:Maria Francisca Xavier)

University of Lisbon (Lisbon): Centro de Linguística da Universidade de Lisboa (CLUL) (C:Amália Mendes); Natural Language and Speech Group (NLX-Group), Department of Informatics (C:António Branco)

University of Azores (Ponta Delgada (Azores)): (C:Luis Mendes Gomes)

University of Porto (Porto): Centro de Linguística (C:Fátima Oliveira); Laboratory of Artificial Intelligence and Computer Science (C:Miguel Filgueiras)

Romania (NCP: Dan Tufiş) Romanian Academy of Sciences (Bucharest): Research Institute for Artificial Intelligence (C:Dan Tufiş)

University Babes-Bolyai (Cluj-Napoca): Faculty of Mathematics and Computer Science (C:Tatar Doina)

"Al. I. Cuza" University of Iaşi (Iaşi): Faculty of Computer Science (C:Dan Cristea)

Romanian Academy of Sciences (Iaşi): Institute of Computer Science (C:Horia-Nicolai Teodorescu)

University of Piteşti (Piteşti): Faculty of Letters (C:Mihaela Mitu)

West University of Timişoara (Timişoara): Faculty of Mathematics and Informatics (C:Viorel Negru)

Serbia (NCP: -) University of Belgrade (Belgrade): Faculty of Mathematics (C:Duško Vitas)

Slovakia (NCP: -) Slovak Academy of Sciences (Bratislava): Ľ. Štúr Institute of Linguistics (C:Radovan Garabík)

Slovenia (NCP: Tomaž Erjavec) Alpineon d.o.o. (Ljubljana): (C:Jerneja Zganec Gros)

Josef Stefan Institute (Ljubljana): Dept. of Knowledge Technologies (C:Tomaž__Erjavec)

Spain (NCP: Núria Bel) University of Alicante (Alicante): Departamento de Lenguajes y Sistemas Informáticos (C:Patricio Martínez-Barco)

Institut d'Estudis Catalans (Barcelona) (C:Joan Soler i Bou)

Technical University of Catalonia (UPC) (Barcelona): Centro de Tecnologías y Aplicaciones del Lenguaje y del Habla (TALP) (C:Asuncion Moreno)

Universitat Autònoma de Barcelona (Barcelona): Facultat de Filosofia i Lletres, Dpt. de Filologia Anglesa i de Germanística (C:Ana Fernández Montraveta)

Universitat de Barcelona (Barcelona): Departament de Lingüística General (C:Irene Castellón)

Universitat Oberta de Catalunya (Barcelona): Department of Languages and Cultures (C:Salvador Climent)

Universitat Pompeu Fabra (Barcelona): Institut Universitari de Lingüística Aplicada (C:Núria Bel)

University of Barcelona (Barcelona): Facultat de Filologia - Ramon Llull Documentation Centre (C:Joana Alvarez)

Autonomous University of Barcelona (Bellaterra): Facultat de Letras, Dept. Filología Española (C:Carlos Subirats)

Girona City Council (Girona): Records Management, Archives and Publications Service (C:Joan Boadas i Raset)

University of Jaén (Jaén): Escuela Politécnica Superior, Departamento de Informática, SINAI group (C:María Teresa Martín Valdivia)

University of the Basque Country (Leioa): Computer Science Faculty, Natural Language Processing Group (C:Arantza Diaz de Ilarraza)

University of Lleida (Lleida): Departament d'Angles i Lingüistica (C:Gloria Vázquez)

Autonomous University of Madrid (Madrid): Laboratorio de Lingüística Informática (C:Manuel Alcantara Plá)

University of Málaga (Málaga): Facultad de Filosofía y Letras, Dept. of English, French, and German Philology (C:Antonio Moreno Ortiz)

Universidad Politécnica de Valencia - ITACA (Valencia): Grid and High Performance Computing Group (C:Vicente Hernández García)

University of Vigo (Vigo): Facultade de Filoloxía e Tradución, Department of English, Research group LVTC (C:Javier Perez-Guerra); Seminario de Lingüística Informática, Departamento de Tradución e Lingüística, TALG Research Group (C:Xavier Gómez Guinovart)

University of Zaragoza (Zaragoza): Facultad de Filosofía y Letras (C:Carmen Pérez-Llantada)

Sweden (NCP: Lars Borin) University of Gothenburg (Gothenburg): Department of Linguistics, Faculty of Arts (C:Anders Eriksson); Språkbanken, Dept. of Swedish Language (C:Lars Borin)

Linköping University (Linköping): Department of Computer and Information Sciences (C:Lars Ahrenberg)

Lund University (Lund): Humanities Laboratory (C:Sven Strömqvist)

KTH Royal Institute of Technology (Stockholm): Department of Speech, Music and Hearing, CSC (C:Rolf Carlson)

Language Council of Sweden (Stockholm): (C:Rickard Domeij)

Swedish Institute of Computer Science AB (Stockholm): (C:Björn Gambäck)

Umeå University (Umeå): HUMlab (C:Patrik Svensson)

Uppsala University (Uppsala): Department of Linguistics and Philosophy (C:Joakim Nivre)

Turkey (NCP: Gülşen Eryiğit) Istanbul Technical University (Istanbul): Elektrics-Electronics Faculty, Computer Science Department, Natural Language Processing Group (C:Gülşen Eryiğit)

Sabanci University (Istanbul): Human Language and Speech Laboratory, Faculty of Engineering and Natural Sciences (C:Kemal Oflazer)

United Kingdom (NCP: Martin Wynne) Bangor University (Bangor): Language Technologies Unit (C:Briony Williams)

University of Birmingham (Birmingham): Department of English (C:Oliver Mason)

University of Surrey (Guildford): Department of Computing, Faculty of Engineering and Physical Science (C:Lee Gillam)

Lancaster University (Lancaster): Department of Linguistics and English Langauge (C:Paul Rayson)

National Centre for Text Mining (Manchester): National Centre for Text Mining (C:Bill Black)

Oxford Text Archive (Oxford): Oxford University Computing Services (C:Martin Wynne)

University of Sheffield (Sheffield): Natural Language Processing group, Department of Computer Science (C:Wim Peters)

University of Wolverhampton (Wolverhampton): Research Institute of Information and Language Processing (C:Constantin Orasan)