CLARIN

A European Research Infrastructure

Newsletter

Number 1, 2008, May

What is CLARIN?

Steven Krauwer
CLARIN coordinator

The easiest way to describe CLARIN in non-technical terms is to say that the goals of the CLARIN enterprise are threefold. First of all, it aims at uniting existing digital archives in Europe that contain language based material into a federation that will allow the social sciences and humanities research communities unified access to the content. Second, it wants to make the wealth of language and speech processing tools that have been developed over the recent years available to interested researchers with a view to opening up new research avenues. Third, it wants to provide web based services that will allow non-expert users (especially humanities and social sciences researchers without a technological background) to perform complex tasks on the materials contained in the archives, such as 'Summarize *Le Monde* of March 17 2008 — in Polish'.

To achieve this a large number of technical challenges will have to be addressed, such as (just to mention a few): how do we unite existing archives from all over Europe into a single federation, how do we solve the problem that many language based resources use many different representation conventions, how do we solve the problem that existing tools and applications each have their own expectations in terms of input and output formats, how do we solve the problem that (in the CLARIN view) all languages are equally important but that the level of technological coverage is quite uneven, how do we make sure that what we will offer in terms of tools and services responds to present and future needs of our primary target audience, how do we protect the intellectual property rights of those who have provided data or tools, and how do we ensure that whatever the infrastructure that we manage to build is sustainable.

Following up on the ESFRI action aimed at identifying potential essential research infrastructures for Europe and the FP7 Infrastructures programme, the EC has funded a number of European Infrastructure initiatives and given them the green light for a preparatory phase during which all of the above issues (and many more) will have to be addressed with a view to actual implementation of the infrastructure. CLARIN is one of them, and in the rest of this newsletter we will describe our approach and expected results towards the end of 2010, when (we hope) we will be able to embark on the next phase: the construction of the CLARIN infrastructure.

The EC funding for this preparatory phase is relatively modest (4.1 million euro) and will just be sufficient to cover the generic, language independent activities in the project. In order to do our work properly we will 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. Unfortunately, our project budget does not allow us to support this participation financially, but, fortunately, all participating countries (at this moment 22) have committed themselves to giving additional support at their national level.

We envisage that at least part of the national CLARIN funding will be used to support participation by additional groups in the project, with a view to ensuring that what is proposed during the preparatory phase will be adequate to serve the interests of the national research communities and languages.

Please contact us if you want to join one of our working groups.

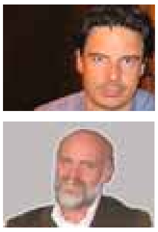
CLARIN
Common Language Resources and Technology Infrastructure

Towards an integrated and interoperable ***research infrastructure*** *of language resources and its technology enabling* ***eHumanities***

Easy access to Language Resources and Technology for the Humanities community

Visit CLARIN web site at http://www.clarin.eu

Editors' Foreword



Marko Tadić
& Dan Cristea
CLARIN Newsletter editors

Dear readers,

In your hands you have the first issue of CLARIN Newsletter, a publication initiated and supported by CLARIN — an acronym for the Common Language Resources and Technology Infrastructure, which is a combination of collaborative projects and coordination and support actions, registered at the EU under the number FRA-2007-2.2.1.2. This publication is planned to appear 4 times per year, both electronically and on paper, at least for the lifetime of the preparatory activities. The electronic version can be accessed by anyone on the project's site at **www.clarin.eu**. On demand, it can be e-mailed to subscribers. The subscription information and forms can be found on the Newsletter section of the CLARIN site. On paper it will be distributed mainly during main language technology and humanities and social sciences events, and only occasionally it will be sent by mail (no budget is put aside for this).

Vision

We want this newsletter to be a publication where people can find the latest news on CLARIN advances and where they can tell their experiences or express opinions. The nice thing about CLARIN is that it facilitates scientific collaboration among people that work in domains which were considered until recently as distant as the Earth's poles. This is the case of humanities and social sciences, on one hand, and computer science, on the other. They will meet in CLARIN for the common enterprise of finding the most appropriate ways in which language resources, whether they are expressed in speech, in text, or multimedia, can be better offered to scientists and to the public at large. In order for this to happen, we invite you to contribute in this Newsletter in two ways: either as a correspondent editor — therefore sending us periodically news from your scientific community, being it your country, your institute or your group, or as an occasional author.

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.

Content

This first issue, of a series which we hope to include 12 of them (the CLARIN preparatory phase extends from January 2008 till December 2010), aims at offering the first contact to scientists with the CLARIN world. You have just turn the page on which Steven Krauwer, the project coordinator, presented the CLARIN mission in the larger context of the ESFRI action. Martin Wynne, who is the port-parole and liaison with the DARIAH (**www.dariah.eu**) project in CLARIN, tries to convince you why a new infrastructure to deal with resources and tools is required by humanists and social scientists. We have dedicated page 4 to two categories of scientists dealing with resources: consumers and producers. In this issue these two voices belong to János László, which brings a motivation for IT needs in social psychology, and Max Silberztein, the developer of the NooJ tool, which was at the very heart of the work reported by János László. The central pages are dedicated to the CLARIN kick-off meeting, hosted by MPI and the Nijmegen City Hall, from 17 to 19 March 2008. Some meetings in which the CLARIN idea took gradually shape are also briefly described. We have thought that many of our readers would like to know in more details the CLARIN structure and how are its activities organized. This is why we have offered the following 4 pages to members of the Executive Board.

First, Peter Wittemburg makes a presentation of the CLARIN organization, and then all coordinators of the CLARIN working packages WP2-WP8, minus WP4 which does not exist, describe their specific activities. The last page, now and in all further issues, is dedicated to a presentation of events, most significant for CLARIN, over a period of 6 months in advance.

Enjoy your reading! **C**

List of national correspondents

Austria
Gerhard Budin

Belgium – Flanders
Inneke Schuurman

Bulgaria
Svetla Koeva

Croatia
Marko Tadić

Czech Republic
Karel Pala

Denmark
Hanne Fersoe

ELRE/ELDA
Victoria Arranz

UK
Martin Wynne

Estonia
Tiit Roosmaa

Finland
Kimmo Koskenniemi

France
William Del Mancino
Bertrand Gaiffe

Germany
Lothar Lemnitzer

Greece
Maria Gavrilidou
Stelios Piperidis

Hungary
Tamás Váradi

Italy
Valeria Quochi

Latvia
Andrejs Vasiljevs

Malta
Mike Rosner

Netherlands
Peter Wittenburg

Norway
Koenraad De Smedt

Protugal
Antonio Branco

Poland
Maciej Piasecki

Romania
Dan Cristea
Dan Tufiș

Spain
Nuria Bel

Sweden
Sven Strömqvist

Language Resources in Humanities and Social Sciences Research

Martin Wynne
CLARIN EB member

Research in the humanities and social sciences is changing. The increasingly widespread use of digital resources is changing the types and the amount of data that we use. The use of electronic data means that we need computational tools to create, annotate, process, store, and analyse them. The ways in which we communicate the frequency with which we communicate are changing too, and so are the ways that we share out data and our research findings.

New requirements for infrastructure

This digital revolution means that the support that is needed to faciliate research is different today. The traditional support infrastructure is no longer sufficient. Desktop computers, a network, email and web access are now the common currency of academic work. Use of electronic datasets requires a further support infrastructure, to allow the transfer and processing of large amounts of data, to allow secure shared access to resources and tools, and to allow effective online collaboration and data sharing. CLARIN aims to build such an infrastructure for the use of electronic language resources in the humanities and social sciences.

> *As well as developing computing infrastructure, CLARIN will be promoting new forms of multi-disciplinary collaborative research for many researchers in the Humanities and Social Sciences...*

One of the most important changes in research today is the increase in collaboration. Digital projects entail collaboration, but many HSS academic communities are more familiar with the model of the lone scholar, researching and publishing alone. As Daniel Pitti notes:

"the most significant impact of information technology may be increased collaboration. Collaboration, when successful, offers many intellectual, professional, and social benefits...collaboration also presents unfamiliar challenges, which require careful attention and time..." (1)

One of the most important challenges for CLARIN is to recognize and cope with the changes in working practices which the increasing use of electronic resources will entail. CLARIN aims to facilitate the deployment of the latest computing tools and infrastructure so that language resources and tools can be used effectively. As well as developing computing infrastructure, CLARIN will be promoting new forms of multi-disciplinary collaborative research for many researchers in the humanities and social sciences.

There are many language resources and tools already in use in research across many subject areas, and there is a clear potential for many more to be successfully developed and deployed. If this potential is to be realised, the developers of the CLARIN infrastructure will have to understand the processes involved in research in the humanities and social sciences, to be sensitive to differences in the cultures of the various subject areas, and to be responsive to user requirements.

The problems which are faced by CLARIN are not new ones. Problems of standards for textual representation, interoperability of tools, and problems with licensing, access and preservation have dogged the humanities since the invention of the digital computer. The language resources which are currently available exhibit a confusing variety of forms of textual representation, metadata, annotation, and access arrangements. Tools are often developed for ad hoc use within a particular project, or for use by one group of researchers, or for use with only one text or set of data, and are not developed sufficiently to be deployed as widely-used and sustainable services. As a result, a large amount of effort has been wasted over many years in many places developing applications with similar functionality. CLARIN offers us an opportunity to address these problems in a systematic, sustainable and global fashion.

Usage of IT in the Humanities

A Summit on Digital Tools in the Humanities in Charlottesville, Virginia in 2006 estimated that only 6% of scholars in the humanities go beyond general purpose information technologies (email, web browsing, word processing, spreadsheets and presentation slide software), and suggested that revolutionary change in humanistic research is possible thanks to computational methods, but that this revolution has not yet occurred (2). This is an exciting time for researchers in the humanities and social sciences, as the introduction of new instruments and datasets make possible new types of research, but it is clear that a new infrastructure is needed for the potential to be realised. C

(1) Pitti, Daniel V. (2005), 'Designing Sustainable Projects and Publications' in A Companion to Digital Humanities, Blackwell: Oxford
(free online at http://www.digitalhumanities.org/ companion/)

(2) Summit on Digital Tools for the Humanities: Report on Summit Accomplishments. 2006. http://www.iath.virginia.edu/dtsummit/SummitText.pdf (Accessed online 25.10.2007)

Users' view

What do social psychologists expect from CLARIN?

János László
Social Psychologist, Hungarian Academy of Sciences

Current research in mainstream social psychology shows that language conveys substantial implicit information about inter-group relations when depicting events. The Linguistic Category Model (Semin and Fiedler, 1988; 1991 and other paradigms all imply that both wording and compositional patterns of the text convey information that function similarly to "hidden curricula" (Jackson, 1968; Snyder, 1970). They can reinforce or alleviate existing stereotypes; shift blame to out-groups or take in-group responsibility for negative outcomes; position in-group as active or passive participant in events.

Our research group, in collaboration with the Institute of Linguistics of the HAS, with the Institute of Informatics of the University of Szeged and with the Morphologic Ltd, aims at uncovering implicit linguistic and narrative devices in history school books, which mediate identification and stereotyping in inter-group relations.

To explore this hidden dimension of language use, we have launched an international project with 12 participating countries involving historians and social psychologists with the specific aim to study systematically the use of linguistic means in inter-group stereotyping in the history school-books in the participating countries. Results will reflect not only the level of elaboration of so called "traditional" regional conflicts, but also tendencies of accepting supranational, e.g. European categories.

Text analysis in both directions (i.e. linguistic means of stereotyping and causal inferencing as well as identity related linguistic markers) will be carried out by computer programs using state-of-the-art language technology tools. The finite-state linguistic development tool, NooJ (www.nooj4nlp.net) is eminently suited to the task on account of its robust lexicon and its highly user-friendly facility to define complex local grammars in the form of finite-state transducers. Collaboration between the Linguistics and the Psychology Institutes of the Hungarian Academy of Sciences have proved very productive and resulted in a set of linguistic categorization algorithms. NooJ is available in all the languages involved in the project, except German. Linguistic categorization algorithms will be transferred to the other languages concerned in the project. Interlingual transfer will be aided by using Wordnet, the hierarchical relational model of the mental lexicon developed at Princeton University (www.wordnet.princeton.edu), which is now available in various other languages as a result of the EuroWordnet, Balkanet and other projects. The Hungarian linguistic categorization modules will be projected onto the InterLingual Index of EuroWordnet, assuring consistency of cross-linguistic transfer. The language-specific local grammars to accommodate the intricate contextual constraints operating in different languages will be worked out under the guidance of the Linguistics Institute of the HAS. The project hopes to draw benefit from the support of the CLARIN. We expect CLARIN to provide us with guidance and support with filling currently lacking language support in the NooJ system.

Editors' note

On this page we will publish opinions, discussions, view and arguments that will usually come from two sides. One will illustrate the standpoint of CLARIN users or "consumers", while the other will try to present the ideas that are coming from the direction of LRT developers.

Developers' view

How can NooJ help?

Max Silberztein
Université de Franche-Comté

It is my pleasure to give an account of a collaboration with social psychologists who are investigating latent psychological concepts in texts. At first blush, tracking down highly abstract concepts like the individual perception of time and studying group or national identity with automated methods of text processing seemed a daunting task.

As it turned out, however, the finite-state linguistic development NooJ has proved a productive tool in the hands of non-specialist colleagues, like the social psychologist team of János László. NooJ is a freeware corpus processor that allows users to apply queries to large texts and corpora. One special feature of NooJ is that queries, which can be very simple or very complex, are based on potentially large linguistic resources such as dictionaries and grammars. NooJ contains large-coverage dictionaries for a dozen languages ranging from French and English to Portuguese, Serbian and Hungarian.

Beside basic sets of linguistic resources, NooJ users can add their own specialized dictionaries and grammars, together with their own system of properties, in order to describe complex sets of terms, expressions and their variants. Then, if a user has constructed a dictionary that contains thousands of technical terms, a basic NooJ query such as <MEDIC> (medical terms) or <MEDIC+PSY> (medical terms of the PSYchological domain) allows users to retrieve, extract and analyze each occurrence of the recognized terms. Moreover, NooJ grammars generalize the notion of terms to handle complex expressions in texts, even when these expressions are discontinuous (such as in "John *took* the problem you mentioned yesterday *into account*").

The possibility of adding resources that are not necessarily "linguistic" has allowed a number of "non-linguist" researchers to use NooJ in a variety of domains. For instance, historians of the Univ. Paris 1 are using NooJ to study how the English legal vocabulary evolved from old French (the language used by the Norman invadors). Psychologists have used NooJ to study how Charcot created a new technical language in order to delimite a new medical domain. Researchers in literature have used NooJ to study how certain concepts (such as "death") are used by certain authors along their lives (e.g. Marcel Proust), etc.

Beside its basic corpus query system, NooJ can be used to analyse and annotate texts. In effect, when applying dictionaries and grammars to a corpus, NooJ annotates it. NooJ contains a number of specific tools to add new lexical, syntactic or semantic annotations to a text, or to remove annotations from the texts' annotation structure (in order to disambiguate texts' analyses). NooJ imports structured XML documents, so that XML tags get converted into "regular" NooJ annotations, and then used in queries or even complex grammars. Reciprocally, NooJ can export all or parts of the texts' annotation system back to XML documents.

The CLARIN Language Resource Ambition

Peter Wittenburg & Tamás Váradi
CLARIN EB members

Nature of Language Resources

For a growing number of researchers, including increasingly in the humanities, analysis of vast amount of data has become the key to successful and empirically based research. The nature of the required resources is very heterogeneous and although linear texts still tend to form the largest amount of data it is structured and highly specialized information extracted from the surface which is most often required in processing language material to answer sophisticated research questions.

Resource Management

The vast majority of this data is currently typically produced by isolated researchers and resides on local computers, unknown and inaccessible to the rest of the community. The individual researchers do not have any digital access and preservation strategy. Data curation and management distracts researchers from the work and imposes requirements they are not properly trained for, a situation that is only aggravated by lack of budget allocated to the tasks.

On the other hand, it is common knowledge now that digital collections need continuous curation and management effort to guarantee long-term accessibility. Data management also includes digital rights management, which tends to be much neglected. Most research projects, departments and institutes face almost insoluble challenges to meet the requirements.

Emerging Business

Currently, we see large activities from interested companies to ensure digital accessibility preferably for those resources where a return of investment can be expected. The commercial goal is certainly to get control of digital content and to establish a business model for trading with all sorts of information, preferably with minimal overhead. Scientific information is just one small slice of the cake. But the companies that have been dealing with printed scientific publications until now have shown that this is a profitable sector.

Recently, Google announced an offer to the research world to store primary and secondary research data (1) — in effect, data that is essential for carrying out modern data-oriented eResearch. In evaluating this initiative the following points should be considered:

1) Judging from widespread commercial practice, we can certainly expect that the current business model offering access free of charge will be changed when a critical mass of content has been reached.

Dennis G. Jerz & Phillip Lenssen: Early draft of how Google home page would look like in 1407 (http://jerz.setonhill.edu/weblog/permalink/google-1407/)

2) The prevalent publishing model in the past centuries has been that the researchers produce and consume research results and that the publishers organize the workflow process including the dissemination. The Open Access Initiative (2) has made it clear, however, that the research community does not readily agree with the way the big publishing monopolies handle information in the electronic age. The current model advocated and practiced by the companies focusing on digital content is a centralized one, i.e. the rules for handling data that is created and consumed in a distributed fashion will be defined centrally by company executive boards. Selection, rating and other important processes will be determined by these companies.

Challenge for CLARIN

We need to understand that the CLARIN research infrastructure is, to a certain extent, in direct competition with proposals such as from Google. In contrast to the model offered by Google and other commerical providers, CLARIN strongly relies on a distributed model fully controlled by the research community. The competition forms an enormous challenge since we need to achieve a high degree of cost efficiency and service quality. However, if CLARIN also wants to guarantee access to the many small languages and the more complex resource types and if CLARIN wants to handle sensitive rights management issues with care, services will be more expensive.

CLARIN — as many other research infrastructures that want to enable the eScience paradigm — needs to be seen as a self-organizing model of dealing with research information in a sensitive way so that it can present a viable alternative to the emerging monopolies.

Emphasizing our strengths is important not just in terms of the competition with commercial companies but also to make an impact on our target research community. The question of incentives is important whatever model we are considering. Why should the individual researchers that we described earlier turn to either Google or to us with their resources or tools?

In order to achieve progress from the currently fragmented situation CLARIN must clearly make itself attractive to the research community and offer easy to use services. Highlighting on the wide ranging expertise that CLARIN will also provide we can demonstrate our usefulness to our target audience and at the same time show our advantage over initiatives driven by commercial interests only.

(1) http://news.bbc.co.uk/2/hi/technology/6425975.stm

(2) http://www.soros.org/openaccess/index.shtml

Kick-off meeting in Nijmegen

Dan Cristea
Editor

CLARIN had its kick-off meeting in the city of Nijmegen, The Netherlands, between 17 and 19 March 2008. The organization was arranged by the Max Planck Institute for Psycholinguistics, with a generous help from the City Hall of Nijmegen.

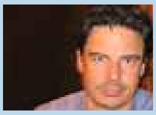
About 80 participants attended the meeting, representing 56 institutions from 23 CLARIN member countries. For most of these people the accommodation was arranged in the small town of Kleve, across the border, in Germany, at about 20 km distance from Nijmegen. The bus brought the participants each morning through the already green country side meadows, but also along a channel and passing a forest in which remains of a frontier check point were still visible.

The first day was devoted to bring to the participants detailed information about all CLARIN aspects so that a broad understanding of the goals could be obtained. It was also meant to give the necessary insight into some details of the work to be realized and to collect comments and recommendations from all CLARIN members. Two consecutive sessions gathered all CLARIN members, independent of whether they are consortium partners of the EC funded project or members of national CLARIN groups. These sessions have been hosted by the City Hall, in the impressive round table

The Round table room of Nijmegen CIty Hall where the plenary sessions took place

The WP5 meeting

Previous meetings



Marko Tadić
Editor

Having in mind the number of different language resources and language technologies activites in Europe during the last decade, it should be clear that CLARIN could not start from the scratch. Moreover, it followed the line of many previous initiatives and projects that, seen from today's perspective, were actually preparing the ground for CLARIN as the project that bring into play the mature technology and establishes it as the essential infrastructure for humanities and social sciences in 21st century.

In the same manner, before the very kick-off meeting in Nijmegen, there has been a series of founding and preparatory meetings where CLARIN was shaped up into the form that it exhibits today.

The first, almost historical one, was in Paris, in February 2006 where initiatives from different research groups such as TELRI and EARL gave the necessary wind in the back to the whole idea. After that meeting where representatives of more than fifteen countries and institutions were present, a series of consultations were organized that led to a common understanding about the general nature of our project.

The second meeting was the workshop organized within the LREC2008 conference in Genoa, in May 2006. This was the first presentation of ideas behind the CLARIN to a wider scientific audience. The time and the place was selected well since that conference is always in the focus of researchers, usually around thousand of them, from the field of language resources and technologies from all around the globe. The series of consultative meetings continued until it was peaked in a finely organized overall CLARIN preparato-

room of the Council of Nijmegen (see photo).

The opening and the first welcome was addressed by the coordinator, Steven Krauwer. Then, from the part of the Municipality, Ms. J. Kunst expressed the honour of hosting this event, and P. Levelt, a former director of the Max Planck Institute for Psycholinguistics, in his speech, evoked the last 30 years since when humanities started to be affected by technology. In his vision of future of humanities CLARIN has one of major roles. This session continued with presentations contributed by the coordinator, explaining the goals and overall structure of the project, by the work package leaders, clarifying the aims, deadlines and milestones in each of their working packages (for WP2 — Peter Wittenburg, for WP3 — Tamás Váradi and Martin Wynne, for WP5 — Erhard Hinrichs, for WP6 — Dan Cristea, for WP7 — Kimmo Koskenniemi and for WP8 — Bente Maegaard), and by the Vice-Chair of the Scientific Board of CLARIN (Dan Tufis), describing the role and activity of different CLARIN boards. To present the content of the Newsletter and its design principles, the WP6 leader invited at the desk Marko Tadić, the editor-in-chief of this CLARIN publication. Draft copies of this first issue were distributed in the audience. Also, Thierry Declerck was asked to shortly present the DFKI experience in constructing help-desk services, which will be exploited and continued in CLARIN, and Dieter Van Uytvanck made a short on-line presentation of the new CLARIN website.

The whole second day was organized in parallel sessions. Participants split, generally following the WP leaders, in seminar rooms to discuss relevant issues in their areas (the picture in the lower left corner shows an aspect of the WP2 meeting). The WP2 and WP5 meetings, touching the very core of the technical aspects of the project, as the former is responsible for building the view on the infrastructure architecture, while the latter should provide the access to language resources and tools, acquired the largest audience and had to be scheduled in two sessions.

The WP3 meeting

The third day was dedicated to the Consortium meeting, addressing the issues of the EC founded project, to working groups (which met as needed), and to the Executive Board meeting. Again hosted in the large City Council room of the City Hall, the Consortium meeting offered the possibility to all WP leaders to resume the previous day sessions and to participants representing the partners to express the state of affairs in their countries. Administrative details, such as consortium agreement, funding, travel possibilities, report writing, deliverable deadlines etc., acquired also the attention of the full audience.

The kick-off meeting made very clear the goals of CLARIN and the rules of the game to all parties involved. In particular, it was stressed to all partners the double funding scheme — national and EC — and to all other members the necessity to urge the formation of their national groups. The attending experts interacted with the working package leaders, the working groups begun to take shape, the activities inside them were initiated, and ideas and suggestions were taken up from everyone to further improve the activity.

The organisers succeeded to create a positive work atmosphere, and, most important, the meeting showed that there is a raising enthusiasm in all countries represented to meet the challenging goals of our project.

ry meeting in Budapest, on 2006-10-30 and 31 where the CLARIN project, more or less as it is shaped now, was founded. The 28 representatives from 21 country were present discussing about the issues that had to be solved in order to make an exemplary project proposal.

The first call for proposals for the Research Infrastructure projects within the FP7 was issued in December 2006 and it was met by the members that formed the consortium of 32 partners between February and April 2007.

Representatives at the Budapest meeting

CLARIN Bodies and Structures

Peter Wittenburg
CLARIN EB member

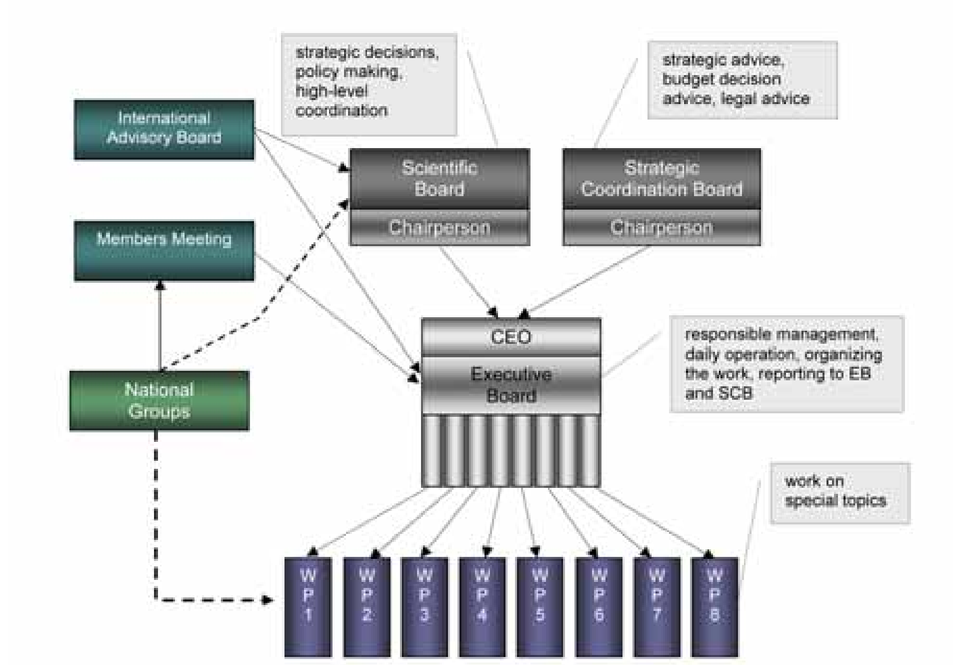
European Research Infrastructures are primarily based on the commitments of the member states. This was clearly expressed in the ESFRI process by the statement that the European Commission will only fund the preparatory phase to facilitate the start of initiatives and to achieve a high degree of synchronization between the member states. While the CLARIN proposal was being prepared, commitment statements from 24 member states were received, indicating great interest in its goals. Currently, some member states have already decided about concrete funding schemes for a number of research infrastructure initiatives, others are busy to decide plotting about national roadmaps and others obviously need more time to decide.

Double scheme

CLARIN will need European and national funds already in the preparation phase to achieve meaningful results. This double scheme is reflected in the organizational setup. At the European level we have the CLARIN project funded by the EC and at the national level we have national groups funded by national funding agencies. The type of organization of the national groups widely depends on national constraints. Similar to the construction of GEANT, the fast European communication network, CLARIN took care in its organizational structure that there is a high degree of interaction between the EC level and the national level activities. The national groups take care that they are represented in the Scientific Board, the Strategic Coordination Board and in the Working Groups that actually carry out the work in the various work packages. Also all CLARIN's registered members, now numbering around 100, have the possibility to directly bring in their wishes and critic via the members meetings.

The Boards

The CLARIN work is essentially planned and carried out by the Executive Board. It contains eight members each of which is responsible for a clearly separable set of tasks clustered in work packages. The Executive Board is fully responsible for the success of the project and needs to report about its plans and decisions to the Scientific Board and the Strategic Coordination Board. The Executive Board has a chairperson who is coordinating the CLARIN work, reporting to the EC and interacting with the different Scientific and Strategic Coordination Board which are controlling the EB work. Each EB member has to take care that the goals will be achieved, i.e. dependent on the requirements and working groups are being formed. The Working Groups can be formed to create specification documents or develop infrastructure frameworks and components.


strategic decisions, policy making, high-level coordination
strategic advice, budget decision advice, legal advice
International Advisory Board
Scientific Board
Chairperson
Strategic Coordination Board
Chairperson
Members Meeting
CEO
Executive Board
responsible management, daily operation, organizing the work, reporting to EB and SCB
National Groups
work on special topics
WP 1
WP 2
WP 3
WP 4
WP 5
WP 6
WP 7
WP 8

Overall structure of CLARIN
(dashed lines indicate that in general members of the SB and working groups will be recruited from the national CLARIN groups)

The International Advisory Board will be formed by international experts and will give advice on all matters to ensure that the CLARIN work will be done in agreement with international trends and that CLARIN specifications have a chance to be accepted by the international community. C

THE IMPORTANCE OF STANDARDS

Laurent Romary
Editor

The creation, maintenance and dissemination of data heavily rely on the conformance to best practices as well as the extensive use of relevant standards. It is all the more essential in order to ensure a high independence between data and tools, too complementary topics that will be dealt with by the CLARIN project. In the domain of linguistic resources, there has been a longstanding implication of the research community to establish standards for the representation and annotation of language structures. Such efforts, carried out in particular within previous EU initiatives such as Eagles or Multext, have progressively been integrated and complemented within projects supported by international standardization bodies or consortia. On the one hand, the TEI (Text Encoding Initiative; **http://www.tei-c.org**) has become the main place where guidelines for the representation of primary textual data are issued and maintained. On the other hand, ISO (International Standard Organization), in continuity with its previous activities on language coding (ISO 639) and terminology representation (ISO 16642), covers the field of language resources within its technical committee TC 37/SC 4, with projects encompassing the various levels of linguistic annotation. As a matter of fact several Lirics partners are strongly involved in the corresponding activities either as coordinators or experts. Besides, one of the essential role of Clarin will be to contribute to the wide dissemination of the corresponding results, and contribute to the establishment of a better interoperability scheme in Europe, while ensuring the these standards accounts for the specificity of our multiple languages. C

WP2: Technical Infrastructure

Peter Wittenburg
WP2 coordinator

One of the core pillars of CLARIN is the technical infrastructure that is needed to overcome the large fragmentation that researchers in particular in the humanities and social sciences suffer from when they want to work with language resources and tools. Few are ready to be used by non-experts and in general it is impossible for them to easily combine resources and tools from different projects to tackle new research questions. A suitable technical infrastructure will rely on a federation of distributed service centers that will offer a wide variety of services to users in a sustainable and unbureaucratic manner. These services range from repository and archiving services for resources and methods for building virtual collections across repositories to web-services that allow more advanced users to build complex applications by combining existing tools.

The basis of all virtual integration and interoperability is a flexible and distributed registry mechanism for all sorts of tools and resources covering for example schemas, corpora, tools as well as ontologies to make them visible and accessible. Federation middleware will guarantee that users can create and access virtual collections by logging in once using their home identity. Unique and persistent identifiers (PIDs) associated with all resources, resource fragments and services will ensure that references will remain stable. In such a federation services will interact based on accepted certificates as a means to prevent misuse.

Services that will help to overcome interoperability problems will range from conversion services that allow users to transform source formats into more generic standardized formats and terminology services that will help to overcome semantic differences, to workflow services that will allow users to combine tools such as taggers and parsers to more powerful workflow engines. Using existing standards and creating new ones where necessary will be required to facilitate interoperability. In this respect CLARIN can rely on previous standardization efforts by initiatives such as EAGLES/ISLE, W3C, TEI and ISO TC37/SC4, and on years of experience in some areas.

To achieve a scalable and distributed infrastructure that also will take care of sub-discipline specialities we will rely on a service oriented architecture and flexible registry mechanisms. With respect to the first we can build on experiences gathered with frameworks such as for example GATE and UIMA for typical NLP applications and LAT for typical tasks when working with smaller languages. With respect to the latter we can refer to the experience gathered with the IMDI, OLAC and DFKI online catalogues. C

WP3: Humanities Projects

Tamás Váradi
WP3 coordinator

Support for humanitites and social sciences is the cornerstone of the CLARIN mission. While language technology and language resources undoubtedly have great potential in facilitating humanities research, this is an area where neither communities have large-scale experience in exploiting the obvious potential benefits. In the light of this situation, we think it is essential that the CLARIN project should establish an active interaction with the research communities in humanities and the social sciences and should gain crucial direct experience about user needs, objectives, data and methods used in research. The most practical means of gaining this experience is through actual collaboration with colleagues in some well-chosen areas.

The purpose of inviting humanities projects to collaborate with CLARIN in the preparatory phase is to enable us to assess the technological, methodological, organizational etc. requirements involved in serving the humanities in the later phases of CLARIN. We are committed to the idea of collaboration with Humanities projects on a prototype scale as the best means of identifying needs and removing any potential obstance from the way of future synergies between the two fields.

To this end, we intend to issue calls to interested humanities projects for collaboration with CLARIN. Unfortunately, the work of the selected projects — we expect to have 3-5 at most — must be funded through independent (i.e. national or other) sources. Our first task will focus on defining carefully the criteria for selecting suitable projects. To the extent that is possible in the limited framework of this work package CLARIN will then provide help such as, for example, advice on how to look for appropriate resources and tools or to convert legacy formats into formats that can be handled seamlessly. The work package will guide the projects to take care that the implementations and results are useful for both the selected Humanities projects and the CLARIN goals.

Another important strand of planned activites related to building links with the humanities communities, exploring the potential stakeholders in the CLARIN mission, rasing awareness and promoting the use of language resources and tools in the humanities fields. To reduce redundancy and optimally exploit available expertise and resources this activity will be carried out in close collaboration with the DARIAH project, which have similar objectives. Executive Board member, Martin Wynne, will be in charge of this area of activities.

Work in WP3 will be coordinated by Tamás Váradi and will involve the following centres: Oxford Text Archive, University of Bergen, Charles University Prague, University of Lancaster, CNR-ILC Pisa, CNRS Nancy and University of Zagreb. C

WP5: Language Resources & Tools Overview

Erhard Hinrichs
WP5 coordinator

This WP deals with specifying, using and implementing *standards for language resources* of all kinds, including e.g. corpora, lexica, grammars and tools for processing them. This is a prerequisite for achieving *interoperability* between linguistic resources and tools. Both will be made available through webservices, and workflows integrating several resources, tools, and services will be defined.

Interoperability of language resources is currently a vital field of research. Several international organisations, including the International Organisation for Standardization (ISO), the World Wide Web Consortium (W3C), and the Text Encoding Initiative (TEI), are active in defining *standards and guidelines* for best-practice in encoding, linguistic annotation, metadata, and webservices. To achieve interoperability we will draw on the use of existing standards defined by these organisations and, where necessary, develop new standards within these relevant standardisation organisations.

Within CLARIN, the term *language resource* subsumes the whole range of linguistic data types such as digitised manuscripts, text, speech and multimodal corpora, lexica, treebanks, typological databases, grammars, ontologies, schemas, and the term *language technology* covers a wide range of processing and annotation components such as taggers, parsers, semantic extractors, manual annotation tools, speech alignment tools, etc. In a first step a *taxonomy* of these heterogeneous types of resources will be defined.

Parallel to the definition of the taxonomy a *survey* of all the existing resources will be initiated. The survey will include a detailed analysis of the structural and encoding characteristics of the resources and the interfaces of the tools that will serve to design a service-oriented architecture. Based on this broad and detailed investigation, a comprehensive taxonomy of language resources and tools will be specified.

In addition to the survey and resource type definition, a *Basic Language Resource Kit* (BLARK) will be specified as a framework for a landscape of language resources across languages. The exact shape of the BLARK has to made be more precise in the course of the project. For a well-documented language, the BLARK might consist of two types of lexica, one form-based and one lexical-semantic, a manually annotated corpus (treebank) and an automatically annotated large-scale corpus, along with standard analysis tools like taggers, parsers and speech recognition systems.

Written and spoken language as well other modalities such as gestures must be included. Potential gaps in the BLARKs of individual languages will be identified by a coordinated action involving the CLARIN members from the involved countries. It will however be left to the national decision boards whether they will fill these gaps.

Another important part of this work package concerns the *implementation* activities as far as they are required in the preparatory phase.

This will focus on the following major tasks: investigating all aspects that have to do with the integration of resources and tools into the infrastructure; studying usage scenarios, including chains of operation in detail and integrating selected language resources into the web service-based infrastructure.

Since we assume that at the national level, several countries will decide to go ahead with language resource and with technology creation and integration, this work package also needs to communicate with the national groups to coordinate these efforts and to determine whether all work is compliant with the standards and practices of the CLARIN infrastructure. C

WP6: Dissemination and advisory services

Dan Cristea
WP6 coordinator

This work package will co-ordinate the generation of content for publicity and the dissemination of CLARIN activities. It will also provide advice and support for researchers in the construction and operational phases of the project. There are three main activities to be taken care of in WP6.

Web site

First, the construction of an online presence where information about CLARIN will be posted. Information should be offered both to the benefit of partners and members of the consortium during the project's lifetime, but also to people outside the consortium. This site is already operational (www.clarin.eu) and is hosted by MPI — Nijmegen. We intend to use it also as the first point of access from where resources and tools organized within the project will be made visible to the whole world.

Newsletter

Second, editing and publication of the *CLARIN Newsletter*, with a periodicity of 4 issues, electronically and/or on paper is a WP6 activity. We hope that the materials of the Newsletter will be contributed not only by the members of the Editorial Board but also by a network of correspondents attached to member institutions from many states in the CLARIN network. Other dissemination materials, such as brochures, leaflets and posters will be also in the responsibility of this work package.

Helpdesk service

Finally, in order to accomplish preparatory work for introducing infrastructure services able to help researchers in the humanities and social sciences and to facilitate new types of research, WP6 will organise a referral helpdesk service. This will have to point interested people to places (centres), projects, experts, documents, web-services or websites where answers to their questions can be found or from where modules can be taken and integrated into their applications.

Based on the experience acquired in similar initiatives (as, for instance, LT World), this task will make proposals for ways to set up, in the construction phase, of a help-desk able to fulfil user needs in a more automatic way, for instance, by inferring what resources and what language processing components could be combined in a processing architecture that will fulfil the desired task on user's primary data. C

WP7: IPR Issues

Kimmo Koskenniemi
WP7 coordinator

This work package will deal with intellectual property rights (IPR) and other legal issues. The IPRs include copyrights for materials and patents for software. In addition, licences and many kinds of agreements for authorization and authentication are necessary for the proper handling and use of language resources and tools.

What kind of copyright licenses we need?

The author of written texts and the speaker of oral works have a copyright and often, a commercial publisher has acquired some of these rights. Language resources are often collected by scholars. The collectors need an explicit license from the author and/or the publisher in order to use and let some other scholars use the resources.

There are problems with existing licenses because there are so many language resources around, created by many authors, published by many publishing companies and collected by many scholars. There may be thousands of licenses which are somewhat different from each other. It is one of the tasks of this work package to study the stock of existing licenses and create a *set of model licenses* to be used when making future agreements between the collectors and the authors/publishers, between the collectors and computing centres, and between the end users and collectors. It is also a task of this work package to find ways for *migrating existing materials* into the framework of new standard CLARIN licenses. There are far more scholars and students than any collector knows directly. It is the task of this work package to find ways (together with WP2) for *managing the licenses and permissions in a reasonable way* to prevent excessive burden to the users (to get a permission) or the collectors (to grant a permission) for the materials but still maintain the confidence of the authors and publishers.

The role of organizations and agencies

Organizations such as ELRA (European Language Resources Association) and ELDA (Evaluation & Language resources Distribution Agency) make language resources available and identify, classify, collect, validate and produce the language resources. The task of this WP is to define the *relation between ELRA/ELDA and CLARIN*, and how they will cooperate.

Open or commercial resources?

Traditionally, research use has been free of charges but this may not be the only alternative. E.g. the Russian Integrum offers almost all published Russian newspapers and periodicals through its commercial on line system which is also used for researchers. The *option of including commercial resources* in the CLARIN scheme will be studied in this WP.

CLARIN needs human language technologies for a wide array of languages as dealt with in WP5 and WP2.

The programs and software modules need to be compatible, i.e. various parts need to be combined to form the services. Software licenses may make this very difficult or impossible in some cases.

This WP will produce recommendations for licenses both for open source and commercial software. C

WP8: Construction and Exploitation Agreement

Bente Mægaard
WP8 coordinator

The problem of building a joint infrastructure has many different dimensions apart from the technological and scientific dimensions addressed in the rest of the CLARIN project. The main task of this activity is to prepare a draft agreement document for national funding agencies to ensure proper funding of the CLARIN infrastructure in the future — after the preparatory phase.

Consequently, all countries participating in the CLARIN project have a representative for this activity. However the participants are all universities or other academic institutions. For the activity to succeed, national agencies will be contacted (with the help of the participants) and will form a working group. A strong interaction with the Strategic Coordination Board and Scientific Board (see page 8) is required.

CLARIN Agreement

We will call the document to be created the *CLARIN Construction and Exploitation Agreement (CCEA)*. The draft CCEA will clearly indicate where complete agreement between the participating countries has been reached and where action needs to be taken at higher political or administrative levels.

European Research Infrastructure

The European Commission is currently developing a legal framework, *European Research Infrastructure — ERI —* for legal entities with the task of running a pan-European research infrastructure. The European Commission has analyzed existing legal structures under national law, such as the Dutch foundations, as well as existing legal forms under Community law, but has found that none of the existing legal forms are appropriate for pan-European research infrastructures.

It is the intention of the Commission to submit a proposal for an ERI framework to the Council as soon as possible. We believe that we can start using the draft proposal in the development of our ideas already as early as the autumn of 2008.

Participation of national agencies

However, before we start discussions with national agencies on their future involvement, we first have to understand the landscape well: how are the national agencies structured in all participating countries, how do they handle research infrastructure, how do they manage funding, in particular funding for international cooperation, what are the known problems, how can we describe best practise, etc. This is the task of the first 12 months.

The fact that the European Commission has started the work on the ERI is a strong support to our work, as this means that all countries will have agreed already to this concept, so our task will be only to adapt this general framework to our specific purposes. C

CLARIN calendar of events

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

May 2008

2008-05-27 to 2008-06-01: LREC2008 Conference with pre- and post-conference workshops, Marrakesh, Morocco

2008-05-27: CLARIN Consortium meeting, LREC2008 Conference, Marrakesh, Morocco

June 2008

2008-06-08 to 2008-06-10: NooJ2008 conference, Budapest, Hungary

2008-06-15 to 2008-06-20: ACL & HLT conference with pre- and post-conference tutorials and workshops, Columbus, Ohio, USA

2008-06-19: PID Services talk, Nijmegen, the Netherlands

2008-06-23 to 2008-06-26: Information Technology Interfaces (ITI2008), Dubrovnik/Cavtat, Croatia

2008-06-25 to 2008-06-29: Digital Humanities Conference, Oulu, Finland

2008-06-26 to 2008-06-28: Stakeholder Workshop "Central and Eastern European Scholarship in the Humanities – harnessing the assets", European Science Fundation, Sofia, Bulgaria

July 2008

2008-07-03 to 2008-07-06: Teaching and Language Corpora Conference, TALC2008 with pre-conference workshops, Lisbon, Portugal

August 2008

2008-08-04 to 2008-08-15: ESSLI 2008, Hamburg, Germany

2008-08-16 to 2008-08-24: COLING conference with pre- and post-conference tutorials and workshops, Manchester, UK

Join CLARIN

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 will 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

Country; Institution; Location; Contact person

Austria: University of Vienna; Vienna; Gerhard Budin

Belgium: ALT (Acquiring Language through technology); Leuven - Kortrijk; Hans Paulussen

Center for Computational Linguistics ; Leuven; Ineke Schuurman

Center for Dutch Language and Speech, University of Antwerp; Antwerp; Walter Daelemans

ELIS-DSSP; Gent; Jean-Pierre Martens

Legal Informatics and Information Retrieval, Katholieke Universiteit Leuven; Leuven; Marie-Francine Moens

Laboratory for Digital Speech and Audio Processing - VUB - ETRO/DSSP ; Brussels; Werner Verhelst

ESAT-PSI/Speech; Leuven; Patrick Wambacq

Bulgaria: Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences; Sofia; Svetla Koeva

Institute for Parallel Processing; Sofia; Kiril Simov

Mathematical Linguistics Departement, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences; Sofia; Ludmila Dimitrova

Croatia: University of Zagreb, Faculty of Humanities and Social Sciences; Zagreb; Marko Tadić

Institute of Croatian Language and Linguistics; Zagreb; Damir Ćavar

Cyprus: Cyprus College / Research Center; Nicosia; Antonis Theocharous

Czech Republic: Charles University; Prague; Eva Hajičová

Faculty of Informatics, Masaryk University ; Brno; Aleš Horák

The Institute of the Czech Language, Czech Academy of Sciences; Prague; Karel Oliva

Denmark: Center for Sprogteknologi, University of Copenhagen; Copenhagen; Bente Maegaard

Dansk Sprognævn - Danish Language Council; Copenhagen; Sabine Kirchmeier-Andersen

Society for Danish Language and Literature; Copenhagen; Jørg Asmussen

Estonia: University of Tartu; Tartu; Tiit Roosmaa

Finland: CSC - the Finnish IT Center for Science ; Espoo; Tero Aalto

University of Helsinki; Helsinki; Kimmo Koskenniemi

Department of Foreign Languages and Translation Studies, University of Joensuu; Joensuu; Jussi Niemi

University of Tampere; Tampere; Eero Sormunen

The Research Institute for the Languages of Finland; Helsinki; Toni Suutari

France: ALTIF; Nancy; Bertrand Gaiffe

TELMA/DIS CNRS; Paris; Florence Clavaud

CNTRL; Nancy; Bertrand Gaiffe

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

Université Paris 4 Sorbonne / CELTA ; Paris; Andre Wlodarczyk

LIF-CNRS ; Marseille; Michael Zock

Germany: Berlin-Brandenburg Academy of Sciences; Berlin; Alexander Geyken

Deutsches Forschungszentrum für Künstliche Intelligentz; Saarbrücken; Thierry Declerck

Institut für Deutsche Sprache; Mannheim; Marc Kupietz

Max Planck Institute for Evolutionary Anthropology; Leipzig; Hans-Joerg Bibiko

University of Frankfurt/Main Comparative Linguistics; Frankfurt/Main; Jost Gippert

University of Leipzig; Leipzig; Codrina Lauth

University of Stuttgart; Stuttgart; Ulrich Heid

Universität Tübingen; Tübingen; Erhard Hinrichs

University of Giessen; Giessen; Henning Lobin

Computational Linguistics Department, University of Heidelberg; Heidelberg; Anette Frank

University of Augsburg ; Augsburg; Ulrike Gut

Greece: Institute for Language and Speech Processing; Athens; Stelios Piperidis

Hungary: Academy of Sciences; Budapest; Tamás Váradi

Budapest University of Technology and Economics Media Research (BME MOKK); Budapest; Peter Halacsy

University of Szeged, Department of Informatics, Human Language Technology Group; Szeged; Dóra Csendes

Iceland: Institute of Linguistics, University of Iceland; Reykjavík; Eiríkur Rögnvaldsson

Icelandic Centre for Language Technology; Reykjavík; Eiríkur Rögnvaldsson

Ireland: National University of Ireland; Galway; Sean Ryder

Israel: Technion-Israel Institute of Technology; Haifa; Alon Itai

Italy: Dipartimento di Linguistica Teorica e Applicata, Università di Pavia; Pavia; Andrea Sansò

Istituto di Linguistica Computazionale; Pisa; Nicoletta Calzolari

Department of Computer Science, University of Rome "Tor Vergata" ; Rome; Fabio Massimo Zanzotto

European Academy Bozen/Bolzano; Bolzano; Andrea Abel

Latvia: Institute of Mathematics and Computer Science, University of Latvia; Riga; Inguna Skadina

Tilde; Riga; Inguna Skadina

Lithuania: Institute of the Lithuanian Language; Vilnius; Daiva Vaisniene

Center of Computational Linguistics, Vytautas Magnus University ; Kaunas; Ruta Marcinkeviciene

Luxembourg: European Language Resources Association (ELRA); Luxembourg; Victoria Arranz

Malta: University of Malta, Dept. of computer science; Malta; Michael Rosner

Netherlands: Meertens Institute; Amsterdam; H.J. Bennis

Data Archiving and Networked Services; Den Haag; Henk Harmsen

University of Twente, Human Media Interaction Group; Enschede; Roeland Ordelman

Center for Language and Cognition; Groningen; Wyke van der Meer

Digital Library for Dutch Literature; Leiden; C.A. Klapwijk

Instituut voor Nederlandse Lexicologie; Leiden; Remco van Veenendaal

Leiden University Centre for Linguistics; Leiden; Jeroen van de Weijer

Centre for Language Studies, Radboud University; Nijmegen; Pieter Muysken

Centre for Language and Speech Technology, Radboud University; Nijmegen; L. Boves / N. Oostdijk

Max-Planck-Institute for Psycholinguistics; Nijmegen; Peter Wittenburg

University of Utrecht/Netherlands Graduate School of Linguistics; Utrecht; Jan Odijk

ILK Research Group ; Tilburg; Antal van den Bosch

Huygens Instituut KNAW ; Den Haag; Karina van Dalen-Oskam

Norway: Dept. of Culture, Language and Information Technology; Bergen; Koenraad de Smedt

Department of Linguistics and Nordic Studies, University of Oslo; Oslo; Janne Bondi Johannessen

Det humanistiske fakultet, Universitetet i Tromsø; Tromsø; Trond Trosterud

Norwegian University of Science and Technology; Trondheim; Torbjørn Nordgård

Poland: University of Wroclaw ; Wroclaw; Adam Pawlowski

Institute of Applied Informatics, Wroclaw University of Technology; Wroclaw; Maciej Piasecki

Institute of Computer Science, Polish Academy of Sciences ; Warsaw; Adam Przepiórkowski

Institute of English Language, Univeristy of Lodz; Lodz; Lukasz Drozdz

Institute of Slavic Studies, Polish Academy of Sciences ; Warsaw; Violetta Koseska-Toszewa

Portugal: University of Lisbon, NLX-Natural Language and Speech Group; Lisbon; António Branco

Romania: Al.I.Cuza; Iasi; Dan Cristea

Institute for Computer Science, Romanian Academy of Sciences; Iasi; Horia-Nicolai Teodorescu

Research Institute for Artificial Intelligence, Romanian Academy of Sciences; Bucharest; Dan Tufis

University Babes-Bolyai; Cluj-Napoca; Doina Tatar

Serbia: Faculty of Mathematics, University of Belgrade; Belgrade; Duško Vitas

Slovenia: Josef Stefan Institute; Ljubljana; Tomaž Erjavec

Alpineon d.o.o. ; Ljubljana; Jerneja Žganec Gros

Spain: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra; Barcelona; Núria Bel

Universitat de Lleida ; Lleida; Gloria Vázquez

TALG Research Group (University of Vigo) ; Vigo; Xavier Gómez Guinovart

Sweden: Lund University; Lund; Sven Strömqvist

Språkbanken, Dept. of Swedish Language, Göteborg University; Gothenburg; Lars Borin

Dept. Speech, Music and Hearing, CSC, KTH ; Stockholm; Kjell Elenius

Uppsala University, Department of Linguistics and Philosophy; Uppsala; Joakim Nivre

Department of Linguistics; Göteborg; Anders Eriksson

Department of Computer and Information Sciences, Linköping University; Linköping; Lars Ahrenberg

Swedish Institute of Computer Science AB ; Stockholm; Björn Gambäck

Language council of Sweden ; Stockholm; Rickard Domeij

HUMlab, Umeå University ; Umeå; Patrik Svensson

Turkey: Sabanci University - Human Language and Speech Laboratory; Istanbul; Kemal Oflazer

UK: Arts and Humanities Data Service; London; Sheila Andersen

Department of Linguistics and English Langauge, Lancaster University; Lancaster; Anna Siewierska

Oxford Text Archive; Oxford; Martin Wynne

University of Sheffield; Sheffield; Wim Peters

University of Surrey; Guildford; Lee Gillam

Research Institute of Information and Language Processing at the University of Wolverhampton ; Wolverhampton; Gina Sutherland

Language Technologies Unit, Bangor University; Bangor; Briony Williams

Department of English, The University of Birmingham ; Birmingham; Oliver Mason

CLARIN newsletter • *publisher* CLARIN • *editor-in-chief* Marko Tadić • *editorial board* Gerhard Budin, Dan Cristea, Koenraad De Smedt, Laurent Romary, Martin Wynne • http://www.clarin.eu