CLARIN Newsletter

A European Research Infrastructure

Number 13, 2011, January-June

CLARIN firmly on the road to become an ERIC

Richard Derksen
Dutch Ministry of Education, Culture and Science, Dept. for Research and Science Policy

First milestone reached...

With the electronic submission by the Dutch Permanent Representation to the EU on 23 May, CLARIN has taken its first step on the road to become Europe's second ERIC.

The submission is really the crown on the hard work done by so many in the Preparatory Phase. The key to this success is of course the high standard of the national CLARIN initiatives but even more importantly the crucial partnership between the national coordinators and their governments resulting in the willingness of these governments to support the set up of the ERIC.

And the support shown by the governments so far has been overwhelming. Fourteen countries and one international organisation have signed the Memorandum of Understanding on the construction of CLARIN-ERIC. The signees are: Austria, Croatia, Czech Republic, Denmark, the Dutch language Union, Estonia, Finland, France, Germany, Greece, Latvia, Lithuania, the Netherlands, Norway and Poland.

....with remarkable speed

The process between the consultation on the signing of the MoU in October 2010 and the submission of the application has been remarkably fast. In January this year it became clear that a sufficient number of countries were willing to send their representatives to The Hague for the First CLARIN-ERIC Steering Committee Meeting. Before this meeting took place the proposed documents for the application, were discussed in an informal meeting with representatives of DG RTD. This already gave a good idea of the elements crucial to a positive decision on the set-up by the Commission. Also the time schedule for the whole set up process was discussed with the Commission.

With the groundwork being done — mostly thanks to Steven Krauwer and Bente Maegaard — the first Steering Committee Meeting took place on 17/18 March 2011. To secure a smooth transition this was a combined meeting of the Strategic Coordination Board and the Steering Committee members plus the national experts. The aim of this meeting was to get as much feedback from the delegates as possible not only on the relevant documents but also on the time frame to have the ERIC set up by the end of 2011. To keep the momentum going the date for the second Steering Committee Meeting was already set for 15 April 2011.

This second meeting everyone understood would be the real test case to reach consensus on the application documents. Main issue for many countries was to get the cash contribution to the ERIC fixed for the first five years. Here it helped that countries prepared to pay an extra annual percentage could commit for less then five years, e.g. four with 25% annual percentage increase. The second issue was the Membership Agreement between the country and the ERIC specifying the in-kind contribution to the ERIC, this actually being the national initiatives. As the national consortia are responsible for these initiatives, many government representatives were not keen on the extra bureaucracy that would result from their ministries signing this Agreement. Looking after the interest of the ERIC itself it was soon clear that what really matters is that the national consortia perform and therefore an Agreement should be made between the ERIC and the national consortia. After this, full consensus was reached swiftly.

What's next

The Commission is now verifying if the application is in compliance with the requirements of the ERIC Regulation. This means for example that DG TAXUD will have a real close look

Continued on the next page ☞

at the way the VAT-exemption is formulated; so it's a good thing that, after some painstaking work, we got their informal approval before we submitted the Statutes.

The Commission will be assisted by independent experts to assess the scientific importance and other matters, such as effective access to the infrastructure and the mobility of knowledge. Here experts who were involved in preparing the ESFRI roadmap will be consulted.

The results of the assessment is expected to follow by the end of June, so in time for the Budapest Meeting of CLARIN National Coordinators on 29/30 June 2011. It should then be clear if modifications of the documents are needed. With the expected positive outcome of the assessment, an invitation to submit a formal request signed by all future members for the set up of CLARIN-ERIC will also come. This means that all countries will be asked by the Netherlands, to sign a letter stating: "Country X hereby requests, through the seat country of the Netherlands, the European Commission to set-up CLARIN-ERIC as an ERIC with the statutes attached hereto". Getting this letter signed, will require a national procedure that varies from country to country. The hope is that, by the end of July, there will be sufficient letters to submit the formal request. Other countries can then send their formal request letters before the Commission decision is signed and still become founding members.

Final steps

On the basis of the formal request, the Commission will prepare its Decision and will ask for the opinion of the ERIC Committee. This Committee is composed of the representatives of all EU Member States. To help them vote favourably, Bente Maegaard and Steven Krauwer will give a presentation at the next Committee Meeting on 1 July 2011. The actual vote for the set up of CLARIN-ERIC is planned to take place at the ERIC Committee Meeting on 21 October of this year.

Following a positive outcome of the vote the Decision will then be signed by Mr. Barosso and published in the Official Journal of the European. The Decision will enter into force on its third day after publication, meaning the ERIC is then officially established. This should be around mid December, so we can have a real good start in 2012. C

Editors' Foreword

Marko Tadić
& Dan Cristea
CLARIN Newsletter editors

Dear readers,

The time has passed as a blink of an eye and we arrived at the end of the preparatory phase of CLARIN. Synchronously with this end, our CLARIN Newsletter comes to an end. This is the last issue — one more than planned, to cover the extenssion of the project until June 2011. It was an extremelly dense period, with a lot of achievements but which has raised also many challenges for the future.

Of central importance these days is the process of establishing the CLARIN ERIC. The front page is telling this story, through the keyboard of Richard Derksen who has been deeply involved in the establishing the first ever ERIC – SHARE. Since the moment this article was written until today, very good news arrived and we felt they should be given to you as they were transferred to the CLARIN community by the coordinator. So, read the Breaking News!

We have invited Steve Brewer to write about the European Grid Infrastructure (EGI) and its connections to CLARIN. It is shown how new services, such as those being developed within CLARIN, can be deployed on resources integrated into the EGI framework.

Two organisations sharing many goals in common, CLARIN and DARIAH, are expected to sign soon a letter of intend for cooperation with EGI, which we reproduce here.

In the middle pages we have an exclusive insight into the fresh policy paper on RI in Humanities from the Standing Committee for the Humanities of the European Science Foundation.

News from countries/regions at the dawn of the CLARIN ERIC are also hosted in this issue. Nederlands is clearly the country which had the most outstanding contribution in the configuration of CLARIN and it's shaping towards a CLARIN ERIC. Jan Odijk brings in his own experience in convincing his government to put up a CLARIN organisation.

Inguna Skadiņa describes LRT and training activities that were developed during the last few years in Latvia (one of the countries that signed the MoU for CLARIN ERIC). The particularly significant implication of Portugal, among all the CLARIN members, in aquiring the objectives of the project is stressed by António Branco and Amália Mendes. Then, Ineke Schuurman explains the work done by CLARIN-Vlaanderen in collaboration with the Netherlands in order to satisfy some of the desiderata of HSS-researchers in Flanders. Finally, Dan Cristea and Gabriela Haja present an important achievement for speakers of Romanian language: the electronic form of the Thesaurus Dictionary of Romanian – an important lexicographic resource, machine readable, that is expected to be handed to SSH researchers very soon.

Enjoy your reading of this Newsletter one last time! C

CLARIN and the European Grid Infrastructure (EGI)

EGI refers to a federated collection of distributed computing resources available to researchers across Europe as well as their international collaborators

Steve Brewer
Chief Community Officer
European Grid Initiative

I am very pleased to be able to take this opportunity discuss the relationship between CLARIN and EGI. I will start with some words about EGI, what it is, what it does and how I see our relationship with the CLARIN community evolving.

As Chief Community Officer at EGI.eu my job is to make sure that this relationship is a success. Whilst we have processes and policies that ensure that EGI's support services and the infrastructure run smoothly and meet the requirements of all our user communities, my role is often to act as a translator and guarantee that messages are successfully passed between researchers, scientists and developers.

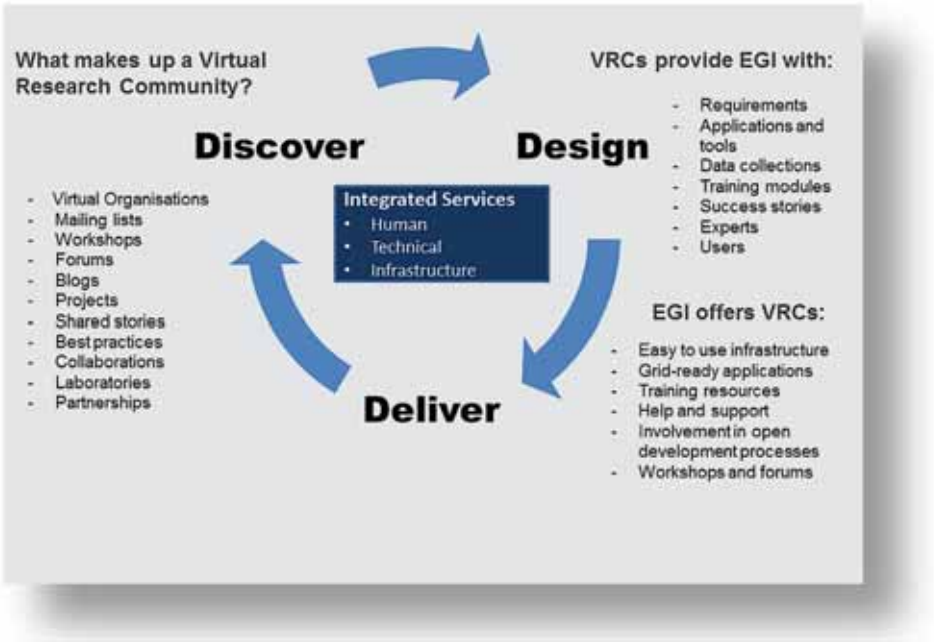
The European Grid Infrastructure refers to a federated collection of distributed computing resources available to researchers across Europe as well as their international collaborators. EGI.eu, on the other hand, is a not-for-profit organisation established to provide coordination for EGI on behalf of the various partners who provide these resources. Typically, these partners are either National Grid Initiatives/Infrastructures (NGIs) or European Intergovernmental Research Organisations (EIROs). It is worth also noting for completeness that EGI-InSPIRE is a part EU-funded, four year project to support the sustainable development of EGI both as an organisation and an infrastructure.

Infrastructure serving many fields

This infrastructure of distributed computing and data storage resources builds on a powerful legacy left by the EGEE project-series which had the added benefit of involving user communities directly within the projects. The development of the grid infrastructure was pioneered by the High Energy Physics community associated with the Large Hadron Collider at CERN. Later, and in conjunction with a number of other communities including Astrophysics, Life and Earth Sciences and Fusion, the infrastructure evolved into a vast and powerful scientific instrument with a wide scope of applications, many tools for submitting computational jobs, managing data and maintaining resources.

This evolutionary process flourished for a decade or so but was ultimately recognised to be unsustainable without a steady flow of significant project funding, and even then was restricted to those research communities that had the inherent skills and inclination to participate in future development.

EGI.eu, the organisation, was thus established to address this specific issue which can be broken down into a number of questions. How do we bring other research communities into the fold? How do we manage the development of the infrastructure to encompass these broader needs? How do we manage the provision of new technology and how do we manage the delivery of all this as an e-Infrastructure with high-quality service, support and ease of use for different levels of user? EGI aims to answer these questions with a model based on a federation of national and international resources providers, supplemented by EGI.eu coordination and support. Responsibility for running the sites and supporting users therefore remains at a national level.



Virtual Research Community cycle

So how does this affect the CLARIN community? The answer is in two ways. Firstly, in terms of infrastructure, EGI now offers a model whereby new services, such as those being developed within CLARIN, can be deployed on resources integrated into the EGI framework and thus benefit from the monitoring and management tools that ensure the smooth and secure running of distributed operations. This allows commu-

Continued on the next page ☞

nities to concentrate on their own areas of specialisation, for example data and metadata structure and model optimisation. The other aspect of such involvement is the benefits of integration within the EGI support framework which I will now describe.

CLARIN within EGI context

Support for users is primarily delivered at a national level, with EGI.eu providing overarching help and support as well as shared services and coordination where necessary. The User Community Support (UCS) team therefore provides a range of integrated services covering infrastructure, technology and human needs. The model for achieving this is the Virtual Research Community (VRC) as we recognise that research communities transcend national borders and clear economies of scale can be achieved by addressing the needs of international communities at the EGI level. VRCs typically comprise more than one Virtual Organisations (VO). However, the VRC concept is adaptable and reflects the fluid nature of any community of practice.

The User Support section of the EGI website (www.egi.eu/user-support) is the place to go to find out more about the tools and applications which are currently available for the infrastructure. As we start to build the relationship between CLARIN and EGI, experts and users from within the CLARIN community will be able to contribute both applications and knowledge which will result in more support resources tailored to the needs of linguists being available. However, the support services that we provide are increasingly being made available through customisable gadgets that can then be embedded directly within CLARIN's own interfaces suitably honed to the particular needs of linguists irrespective of their own computing skills.

Shared vision

The CLARIN project has set itself the goal of establishing a large-scale infrastructure of language resources for the whole European Humanities and Social Sciences community. I welcome the opportunity to share in this vision and I am certain that the framework and services of EGI could make a significant contribution to helping scholars to make the most of the distributed and integrated resources that will emerge from this process.

CLARIN

DARIAH

Digital Research Infrastructure for the Arts and Humanities

EGI

Letter of Intent for Cooperation

Emerging infrastructures in the arts, humanities and social sciences build largely on digital technologies. ESFRI and the European Commission (EC) support the cooperation of emerging digital infrastructures, particularly in the management of scientific information as outlined by the "Riding the Wave" report of the EC's High Level Expert Group on Scientific Data[1].

This letter underlines the intention of the signees to ensure that their infrastructures interact, to share their agendas in infrastructure development, and to continually explore opportunities for collaboration. The prime scope of this collaboration is to research, construct and operate infrastructure services for the arts and humanities.

Some of the topics for discussion may include the following. As one mechanism for ensuring ongoing interaction, the parties will explore and test the creation of Virtual Research Communities (VRC) through EGI — either through multiple channels (e.g. per discipline in the arts and humanities) or as a single shared VRC.

- long-term data storage (i.e. bit preservation) and availability as live data (e.g. streaming of large images and videos, transaction management and version control)
- community annotation and curation of datasets (and thereofore e.g. authority and version control, provenance tracking)
- multiple entry-points suiting the interfaces of virtual research environments (e.g. based on WebDAV or Cloud-like API's)
- hosting and monitoring of middleware services
- simple AAI for large (and potentially unbounded) user groups
- web service frameworks supporting features including secure access control, user-controlled interactive workflows, and computation near data
- connections to shared services in other countries and sectors (e.g. museums and libraries, commercial data providers, domain-specific services)

The signees constitute both key infrastructure as well as arts and humanities communities in Europe; however, they recognise that this collaboration will benefit greatly and, indeed, will only be sustainable through including other partners, including similar initiatives outside of Europe (e.g. Bamboo in the USA).

to be signed by

Steven Krauwer for CLARIN

Laurent Romary for DARIAH

Steve Brewer for EGI

Notes

1 Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data (European Commission). October 2010. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

2 Virtual Research Communities (VRC) in the European Grid Infrastructure (EGI). https://documents.egi.eu/public/RetrieveFile?docid=253&version=1&filename=VRC_accreditation_v1.0.2.pdf

CLARIN NL Story
How to sucessfully convince your government

Jan Odijk
Utrecht University

One of the main goals of the CLARIN-EU preparatory project was convincing the national governments to participate in CLARIN and make available funds for designing, implementing and exploiting the CLARIN infrastructure.

In the Netherlands we were so fortunate to convince the government and to obtain an early and positive decision on a significant budget for the Dutch national CLARIN project. The CLARIN-NL project started in April 2009, so there was a significant overlap with the CLARIN-EU preparatory project. This enabled many researchers from the Netherlands to contribute actively to the CLARIN endeavor, both on the national and on the European level. In the Netherlands we started up a range of projects, which can be categorized in three classes: technical infrastructure projects, demonstrator projects, and data curation projects. In addition, we organized workshops and tutorials on infrastructural matters such as metadata and data categories (ISOCAT), which were often also attended by researchers from outside the Netherlands.

In the technical infrastructure projects, the candidate CLARIN centres in the Netherlands are working together to design and implement the technical infrastructure, continuing the work started in the CLARIN-EU preparatory project. In this way each of them can become a recognized CLARIN centre and they together will form the backbone of the Netherlands part of the CLARIN infrastructure. In the demonstrator and data curation projects existing tools and data are being made CLARIN-compliant, so that they can populate the CLARIN infrastructure and seamlessly 'interoperate' with other data and tools. We paid special attention to reaching the intended user group (esp. Humanities in the Netherlands) and have projects in the areas of history and political sciences, literary studies, archeology, media studies, and of course various branches of linguistics (theoretical, historical, descriptive, language acquisition, sign language, lexicology, typology, and others)

A few projects deserve special mention. First, CLARIN-NL was able to financially support the CKCC project on circulation of knowledge and learned practices in the 17th century Dutch Republic, which was judged in the evaluation of the projects submitted to the CLARIN-EU Call for Collaboration with Humanities and Social Sciences Projects as having "the potential of demonstrating the widespread deployment of the CLARIN infrastructure in a convincing manner" (CLARIN Newsletter 6, p. 4). Second, we set up, together with Flanders, a cooperation project in order to maximize synergy on the shared Dutch language, in which (inter alia) a range of existing HLT tools for Dutch are being converted into CLARIN-compliant web services in a work flow system.

The CLARIN-EU preparatory project is now about to finish. The formal application for the CLARIN ERIC, which the Netherlands will host, has been submitted, and we are confident that CLARIN can become a success. National activities in other countries have also started up or are about to start, and we look forward to working together again with all CLARIN partners, with which we have had a very fruitful cooperation (and fun!) in the preparatory project when working on the CLARIN infrastructure.

The building of the Ministry of Education, Culture and Science of the Kingdom of Netherlands

Research Infrastructures in Digital Humanities revisited

Standing Committee for the Humanities of the European Science Foundation produced the Policy Paper that sheds the new light on the role of Research Infrastructures in the Humanities. It is not published yet, but we exclusively got permission to present its contents to our readers

Marko Tadić
Croatian representative in the ESF Standing Committee for the Humanities

In 2009 the Standing Committee for the Humanities (SCH) of the European Science Foundation (ESF) established a working group on Research Infrastructures in the field of Humanities. This group was initiated by the SCH president Milena Žic Fuchs following the pressing need to get an overall view of RI in the Humanities and to propose relevant directions and guidelines for their usage. The task of this group was to investigate, discuss and come up with an ESF policy paper on RI in the Humanities. The policy paper *Research Infrastructures in the Digital Humanities* represents the current state of views and positions that SCH will support and spread regarding RI. The members of the group were selected from different Humanities disciplines in order to reflect various needs that particular disciplines express towards RI. The members are Maria Ågren (Department of History, Uppsala University, Sweden); Andrea Bozzi (Istituto Linguistica Computazionale, CNR, Italy); Margaret Kelleher (An Foras Feasa, National University of Ireland Maynooth,

Steven Krauwer presenting the CLARIN case study

Ireland); Kristin Kuutma (Institute of Cultural Research and Fine Arts, University of Tartu, Estonia); Claudine Moulin (Fachbereich II / Germanistik, Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften, Universität Trier, Germany, chair of editorial group); Marko Tadić (Department of Linguistics, Faculty of Humanities and Social Sciences, University of Zagreb, Croatia).

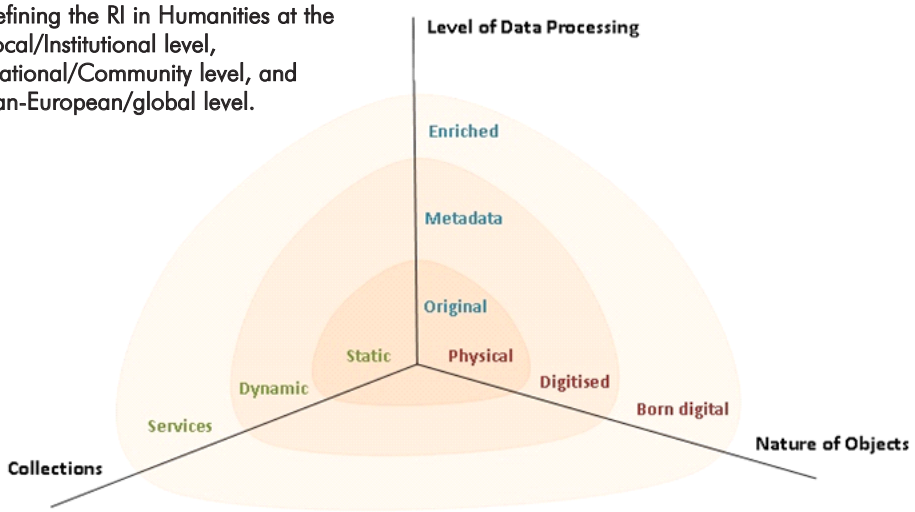
In the process of writing this policy paper the group received help from Arianna Ciula (Humanities and Social Science Unit, European Science Foundation, France) and Julianne Nyhan (Fachbereich II / Germanistik, Kompetenzzentrum für elektronische Erschliessungs- und Publikationsverfahren in den Geisteswissenschaften, Universität Trier, Germany).

The policy paper is composed of several chapters and we will list here their titles to give you the overall picture of its contents: Definitions, taxonomies and typologies of RI; Bridging physical RIs in the Humanities with digital RIs; Researchers input and engagement in producing RIs; Digital research in the Humanities: who is responsible for RIs?; Preservation and sustainability; Evaluation of digital research and its outputs; Communites and practice; Linguistic variety and transnational RIs; Education and training; Priorities and future research Directions.

The input for this policy paper was also collected during the *Strategic Workshop on research communities and research infrastructures in the Humanities*, that was organised by ESF SCH in Strasbourg, 29-30 October 2010 and where different digital humanities projects participated. Among other experts, the coordinator of CLARIN, Steven Krauwer was invited speaker and there he presented the CLARIN project in its current preparatory phase. The DARIAH coordinator Peter Doorn was also one of the invited speakers so the two most important ESFRI projects in the Humanities were fully represented. Beside the chapters listed before, the policy paper brings sixteen different case studies that illustrate the variety of flavours that RI may have in different Humanities disciplines.

This policy paper depicts the overall framework for CLARIN project and CLARIN ERIC as well, but it is also a result of input from several CLARIN members either as direct authors to larger or smaller extent, or as reviewers of the whole paper. We are looking forward to its final publishing.

A set of concurrent criteria for defining the RI in Humanities at the Local/Institutional level, National/Community level, and Pan-European/global level.



On definitions of RI

Special attention in the paper was given to existing definitions or RI and set of criteria that are needed for defining the RI. Several existing definitions are quoted such as the one in the ERIC regulation:

> "research infrastructure" means facilities, resources and related services that are used by the scientific community to conduct top-level research in their respective fields and covers major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives or structures for scientific information; enabling Information and Communications Technology-based infrastructures such as Grid, computing, software and communication, or any other entity of a unique nature essential to achieve excellence in research. Such infrastructures may be "single-sited" or "distributed" (an organised network of resources).[1]

or the definition proposed in the MERIL project:[2]

> A European Research Infrastructure (RI) is an existing facility or (virtual) platform that provides the scientific community with resources and services to conduct top-level research in their respective fields. These RIs can be single-sited or distributed, and can be part of a national or international network of facilities, or interconnected scientific instrument networks. The infrastructure should:
>
> - offer top quality scientific and technological performance and training, that should be recognised as being of European relevance;
> - offer open access to all scientific users on the basis of excellence;
> - have stable and effective management;
> - have the capacity to generate socio-economic impacts.

These differing definitions clearly show that definitions that the concept of research infrastructures can be adapted for and by different disciplines.

[1] Council Regulation (EC) No 723/2009 of 25 June 2009 on the Community legal framework for a European Research Infrastructure Consortium (ERIC), OJ L 206, 8. 8. 2009, p.1.

[6] Mapping of the European Research Infrastructure Landscape; for more details see http://www.esf.org/meril.

CLARIN in Latvia: Conclusions and Vision of Future

Inguna Skadiņa
Institute of Mathematics and Computer Science, University of Latvia

Latvia is among countries that joined CLARIN project consortium after official beginning of the project. Thus activities of the Institute of Mathematics and Computers Science of University of Latvia (IMCS UL) during CLARIN preparatory phase were supported by the Ministry of Education and Science of the Republic of Latvia. This financial support was mainly granted for participation at FP7 CLARIN project, not so much for national activities, e.g. creation of language resources and tools.

During the preparatory phase initial network of CLARIN users and creators of language resources and technologies in Latvia was established, Latvian language resources and tools were identified and first Latvian BLARK was created, several Latvian language tools were adapted to CLARIN standards, practical workshops for humanities users and national workshops were organized. Thus by the end of CLARIN preparatory phase the basic elements for research infrastructure of language resources and technology are established in Latvia.

Everita Andronova presents corpus of modern Latvian at National workshop

Language Technologies and Resources

As in most countries involved in CLARIN project we started with resource and tools inventory. 31 language resource and 11 tools from Latvia are registered at CLARIN Resources and Tools Repository and at CLARIN Latvia Website (www.clarin.lv). Most of registered resources and tools are in proprietary formats of their creators. During preparatory phase IMCS adapted several tools to ISO family standards. However standardization of LRT is still at the initial stage and needs significant contribution during construction phase.

During preparatory phase experimental identity federation LAIFE, based on information system LUIS by University of Latvia, has been set up. The LUIS identity federation is used not only by the University of Latvia but also by most of the regional universities. Thus technically identity federation can be used already now.

The CLARIN BLARK table revealed many gaps of LRT for Latvian, including lack of National corpus, necessity for annotated data and basic language processing tools, e.g., parser, NER and others. Also work on speech resources and technologies needs to be intensified.

As there are many urgent needs related to creation of LRT, it is important to start Latvian Language Technology Program through which work on filling BLARK gaps will be supported and coordinated.

Training

The CLARIN infrastructure aims to serve different communities of linguists and the humanities scholars. However, in Latvia many potential users of CLARIN infrastructure are not familiar with language resources and technologies, thus practical workshops explaining corpus use in research, teaching how to annotate written texts and speech data were organized. The main audience of these workshops were university teachers, researchers and students from master's and doctoral programmes. These practical workshops were well attended and revealed great interest from potential users as well as gaps in education.

Moreover, national workshops were organized to inform CLARIN network not only about CLARIN activities and progress, but also to inform about related national and international activities.

National Advisory Board

To prioritize goals and tasks of the CLARIN initiative in Latvia and to facilitate creation of the CLARIN infrastructure, the CLARIN National Advisory Board was established and approved by the Ministry of Education of Science. The Advisory Board consists of 17 members from universities, research institutes, government organizations and enterprises. Tasks of the Advisory Board include setting priorities and providing recommendations related to the goals of CLARIN project. It is the central body to coordinate work of CLARIN network in Latvia.

CLARIN ERIC

CLARIN is important initiative not only for humanities but for the whole academic society of Latvia. Since the beginning of the CLARIN initiative Latvia desired to participate in CLARIN and contributed actively to the aims of the preparatory phase. CLARIN Latvia activities were regularly supported by the Ministry of Education and Science not only financially, but also in research community and at political level in government. The advancement of CLARIN is mentioned in Action Plan for Implementation of Guidelines for Science and Technology Development accepted by the Cabinet of Ministers in 2010. The Ministry of Education and Science supports creation of CLARIN ERIC and has signed CLARIN-ERIC Memorandum of Understanding. Now the Ministry actively participates in meetings of CLARIN ERIC Steering Committee aiming at creation of CLARIN ERIC.

The Portuguese language in the CLARIN initiative

A snapshot of the first three years

António Branco
University of Lisbon, Department of Informatics

Amália Mendes
University of Lisbon, Center of Linguistics

Launched as a project for preparative work towards the implementation of an European Research Infrastructure, the CLARIN project had a profound impact in the landscape of institutions working on language resources and technology for the Portuguese language that, we may say it now, went well beyond the most optimistic plans or expectations.

The formation of the CLARIN-PT network had a tremendous leverage effect in the gathering of research institutions, centers and groups that otherwise were pursuing their common goals in a very scattered fashion. This resulted in what proportionally (e.g. number of inhabitants, number of R&D centers in the country, national expenditure in R&D, national GDP, etc.) is the largest and the most multi-disciplinary national CLARIN network in our project. So far, CLARIN-PT includes 17 centers, ranking third in absolute figures for constituency among all 33 national networks. These centers are located all over the Portuguese territory, in the major towns of Lisbon, Porto, Braga, Coimbra, ,Covilhã, Évora, and Ponta Delgada. They are working on natural language and speech science and technology and are coming from the areas of Language Studies, Linguistics, Psychology, Computation and Electrical Engineering.

The momentum raised by the Portuguese experience in setting up its CLARIN national network was inspiring enough to make it jump beyond European borders. Despite Brazil not being a EU member or one of the associate states that could benefit from the support of European R&D policies, the fact that there is a very active and tight international community working on the Portuguese language helped to set in motion the formation of a national CLARIN network in Brazil. This network is now in place and eagerly looking forward to intensifying the cooperation with and to better anchoring into the future CLARIN research infrastructure.

The importance of these CLARIN activities and what they represent in terms of the strength and international projection of our community working on the Portuguese language were one of the most important factors in the last few years to promote the visibility of our research area to the Portuguese authorities and policy decision makers. At the closing of the three year preparatory phase of CLARIN, this leaves us very confident that this initiative will be a success and that CLARIN-PT may be a founding member of our future research infrastructure.

CLARIN-Vlaanderen

Ineke Schuurman
Centre for Computational Linguistics (CCL), Katholieke Universiteit Leuven

In the past three years, CLARIN-Vlaanderen has succeeded in stirring enthusiasm for the use of computational tools and resources by HSS-researchers, especially junior researchers. Some more established senior researchers were somewhat more reluctant to make the switch. This was especially the case when their discipline dealt with older forms of Dutch: most of the tools and resources we currently have to offer concern modern Dutch. Furthermore, most texts in older forms of Dutch, especially manuscripts, are not yet available in machine readable format.

Therefore CLARIN-Vlaanderen has initiated workshops concerning older texts: the DigiHist workshops which are to result in a set of digital tools for paleographic materials and the encoding of manuscripts in a machine readable format. This toolkit is expected to be useful for other languages as well. A second series of workshops, called GALATEA I/III, are devoted to promoting what is already available for older texts, and encouraging NLP-researchers to solve smaller problems in this area.

For modern Dutch, several projects were launched, in collaboration with the Netherlands, in order to satisfy some of the desiderata of HSS-researchers (concerning stylometry, use of treebanks, transcription of speech, etc.) and especially a large pilot study, TTNWW, in which a series of tools used in STEVIN projects have been adapted to a web service-based workflow; and annotation formats which over the years have become de facto standards for Dutch are confronted with international standards and best practices promoted by CLARIN, all concepts are defined in ISOcat etc. Structural problems uncovered in this project will be reported back to the international CLARIN community.

As far as the future is concerned, Flanders intends to join ERIC via the NTU (Dutch Language Union), an international Flemish-Dutch organization. This is because Flanders, being a region instead of a country, can not become a member of ERIC itself. Joining the ERIC will give Flemish researchers access to the wealth of resources, tools and services available within the CLARIN-community all over Europe.

Also in the next phases, CLARIN-NL and CLARIN-Vlaanderen intend to join forces, in order to minimize the duplication of effort. Together we will serve the CLARIN HSS-community sharing an interest in the Dutch language, both in its modern and older forms.

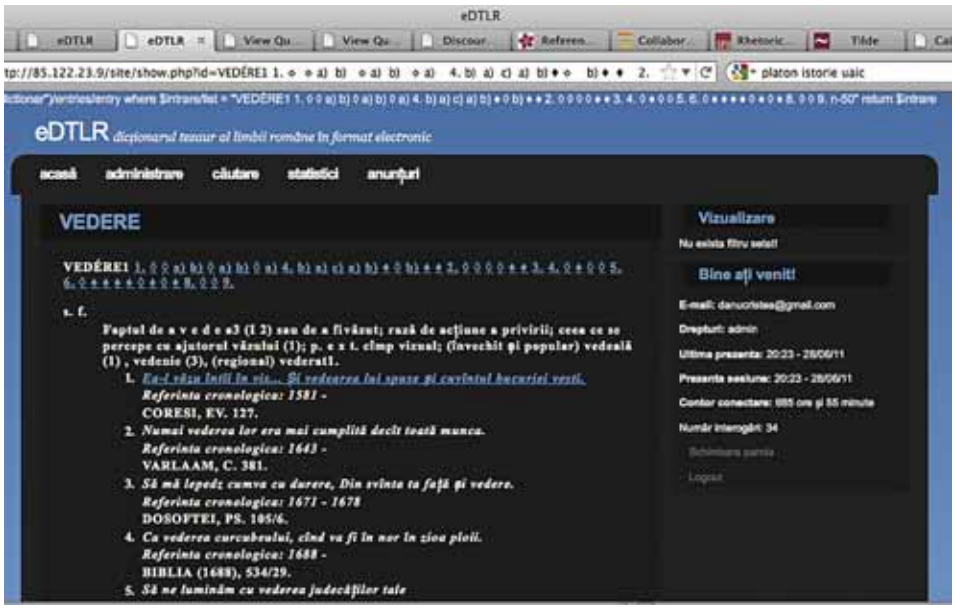

Dan Cristea
Faculty of Computer Science, Alexandru Ioan Cuza University of Iaşi

Gabriela Haja
"A. Philippide" Institute of Romanian Philology, Romanian Academy

The

The dictionary we describe in this article is the Thesaurus Dictionary of the Romanian Language, the biggest lexicographic resource for Romanian language, built and published on paper by the Romanian Academy between 1913 and 2010.

The Web gives access nowadays to extremely many digital and on-line dictionaries. Some of them are intended to be used both by humans and machines. We could say that the dictionaries are moving from paper onto the Web. Here are some notorious examples: the Oxford Advanced Learner's Dictionary, Collins Word Exchange, Merriam-Webster dictionaries for American English, the famous Trésor de la Langue Française informatisé (one of the largest on-line dictionaries of the Romance languages: 100,000 words, 270,000 definitions, and 430,000 examples), Tesoro della Lingua Italiana delle origini, Diccionario de la Lengua Española, etc.



The interface to the Thesaurus Dictinary of the Romaian Language

Roots

The process of building and publishing the Thesaurus Dictionary of the Romanian Language took almost one century. The old series, known as the Dictionary of the Academy (DA) includes 44,890 entries and has been developed between 1913 and 1947. After an interruption, the work was restarted in the middle of the 7th decade of the last century with the new series, known as the Dictionary of Romanian Language (DLR).

The last volume was finally published by the Editing House of the Romanian Academy at the beginning of 2010. In all, DA and DLR have 36 volumes, more than 15,000 pages, about 175,000 entries and approximately 1,300,000 examples. The dictionary was created in the traditional pencil-and-paper way, with citations collected from more than 4,000 volumes of the written Romanian literature.

eDTLR is the name of the digital form of DA+DLR, including its sources in digital form and the software to access them.

The technology

The eDTLR project begun with scanning the printed dictionary volumes which have not been published by using electronic technology (as happened with only the very last few). The images thus obtained, each including two A4 pages, were split, de-skewed, cleared of black margins and downscaled from 600 dpi to 300 dpi, in preparation for Optical Character Recognition. Apart from the different print quality of the paper version, which varies considerably among the volumes published over such a long period of time, a supplementary difficulty in OCR was introduced by the multiple alphabets used in the dictionary: Latin, Greek, and Cyrillic.

In parallel we have started to process the huge number of bibliographical sources of the Dictionary. More than half of the sources have been scanned and OCR-ed, verified for page sequences and stored on the eDTLR server.

In order to remove the errors introduced by the OCR on the dictionary files, we have decided for a combined novice-expert two steps proofreading process. A Web-portal was open to all those willing to contribute — which proved to be a large community of volunteers.

The HTML/Word format saved by the proofreading interface was cleaned by formatting garbage and brought to a lexicographic standard (XML-TEI) by a parser (2). The parser implemented a three steps process. In the first step a configuration of markers is used to identify the different types of fields in the entries. The sense tree of each entry is then determined by exploiting another level of markers and, finally, the atomic definitions (fine-grained senses) are extracted by means of a third level of markers. As such, the parser first identifies the sense markers in a breadth-first manner, and only afterwards builds the sense tree. This strategy, although resembling other known approaches for dictionary parsing (for instance the one used by the French TLFI), brings as a novelty the hierarchical arrangement of the markers, declaratively separated from the parser code. The structures thus obtained have a high degree of

Thesaurus Dictionary of Romanian Language *in Electronic Form*

accuracy, but if the parser leaves behind small unparsed regions, due to errors in the input format, they are being now corrected by lexicographers.

The last phase of the project has concentrated in building links between the citations in eDTLR's entries and their corresponding sources. At a mouse click on a citation, a context, displaying a segment of the original page from where it has been extracted, appears. This context could be larger or shorter, in conformity with the limits allowed by the intellectual property rights regulations of the source.

Conclusions

Benefits of such a large digital dictionary go towards easiness of access, large dissemination for speakers of Romanian, benefits for natural language processing and, not the least, a dramatic change in the manner in which lexicographers' work will be pursued from now on. Also, not negligible, the Dictionary in this form can be published cheaper, while also providing sophisticated indexes between word occurrences, including links to occurrences outside the dictionary itself, in other linguistic thesauri of Romanian or even in other languages.

Acknowledgements

This research described here was supported by the grant no. 91_013/18.09.2007 of the Romanian Ministry of Education, Research and Youth. We thank to all our collaborators, as follows: the Faculty of Computer Science of the Alexandru Ioan Cuza University of Iaşi (coordinator), the Institute of Linguistics "Iorgu Iordan — Alexandru Rosetti" of the Romanian Academy — Bucharest, the Institute of Romanian Philology "Alexandru Philippide" of the Romanian Academy — Iaşi, the Institute of Literary History "Sextil Puşcariu" of the Romanian Academy — Cluj-Napoca, the Research Institute of Artificial Intelligence of the Romanian Academy — Bucharest, the Research Institute for Computer Science of the Romanian Academy — Iaşi, and the Faculty of Letters of the Alexandru Ioan Cuza University of Iaşi. We are grateful also to all our collaborators, professors, students and the public at large, in Romania and the Republic of Moldova, who have contributed to the first proofreading phase in the elaboration of eDTLR.

References

1 Cristea, D., Forascu, C., Raschip, M., Zock, M. (2008). How to Evaluate and Raise the Quality in a Collaborative Lexicographic Approach, Proceedings of LREC-2008, Marrakech.

2 Curteanu, N., Moruz, A., Trandabăţ, D., Extracting Sense Trees from the Romanian Thesaurus by Sense Segmentation & Dependency Parsing, Proceedings of CogAlex Cognitive Aspects of the Lexicon: Enhancing the Structure, Indexes and Entry Points of Electronic Dictionaries, COLING 2008, pp. 55-63, ISBN 978-1-905593-56-9.

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 Ćavar)
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 (ALTIF) (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 Electronique 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)

The participants of the last CLARIN meeting in the preparatory phase, Budapest, 2011-06-29

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 (Reykjavík): (C:Eiríkur Rögnvaldsson)
University of Iceland (Reykjavík): Institute of Linguistics (C:Eiríkur 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 Skadiņa) Tilde Language Technologies (Riga): Tilde Language Technologies (C:Andrejs Vasiļjevs)
University of Latvia (Riga): Institute of Mathematics and Computer Science (C:Inguna Skadiņa)
Lithuania (NCP: Ruta Marcinkevičienė) Vytautas Magnus University (Kaunas): Center of Computational Linguistics (C:Ruta Marcinkevičienė)
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:Torbjoerg 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 Wrocław (Wrocław): Instytut Informatyki Stosowanej (C:Maciej Piasecki)
Wrocław University of Technology (Wrocław): 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:Asunción 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): Facultad 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 (Leida): Departament d'Angles i Linguistica (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)

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