

Fostering Language Resources Network

<http://www.flarenet.eu>



A new European Network for Language Resources – Perspectives for International Cooperation

Nicoletta Calzolari (coord.)
glottolo@ilc.cnr.it



Old Slide: 2004

Political & Strategic Issues

Need of ... International research infrastructures for LRs

- How coordinate national initiatives (ERA)
- No HLT action line → difficulty in designing common global long-term strategies (more opportunistic)



**Need of bodies where to discuss
a broad research agenda & strategic actions
for W&S LRs together**

Technical scientific issues are obviously important

BUT organisational, coordination, political issues play a major role

SO ... should we organise ourselves???



Today

Why a Network of LRs & LTs?

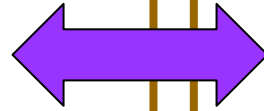
Many dimensions around the notion of language

We need to consider together

- **technical,**
- **organisational,**
- **strategic,**
- **economic,**
- **cultural,**
- **political** issues wrt LRs & LTs

finally

A new EU Networking Initiative: FLaReNet



Need of bodies for a broad research agenda & strategic actions for LT&LRs (W/S /MM) based on all the dimensions

Multilingualism

Inter- & Multidisciplinarity

Technical, scientific issues

Cultural issues

→ Language ... and **cultural identity**

Economic, social issues

- Applications
- Services

Political issues

e.g. a commonly agreed list of requirements for “national” LRs

Sensitive



Today, many vitality & success signs



- In Spoken, Written, Multimodal areas in new emerging areas
 - Different dimensions & layers: Content (Ontologies), Emotion, Time, Cognition, ...
 - Statistical, data-driven approaches
 - Robustness, coverage, ...
 - For Training
 - For Evaluation
 - ...
-
- **LREC** (> 900 submissions; > 1100 participants)
 - Many LRs at **COLING** and even at **ACL**!!
 - **ELRA** (self-sustaining) & **LDC & Jap & Chin**
 - **LRE** (new Journal: N. Ide & NC)
 - **ISO-TC37-SC4/WG4** (International Standards for LRs)
 - **AFNLP...**
 - **New calls or initiatives** in EU, US, ASIA, on LRs, interoperability, cooperation, ...
 - **ESFRI - CLARIN**
 - **→ FLaReNet**



Today

The wealth of data & of basic technologies is such that:

➡ We should reflect again at the field as a whole & ask if

❑ Standards

← Content interoperability

❑ Creation of LRs

← Collaborative creation & Manag.

❑ Automatic acquisition

← Dynamic LRs

❑ Distribution

← Sharing

are still “the” important components,
or how they have changed/must change

... Which new challenges/initiatives towards a
new & more mature infrastructure of LRs & LTs??



Many LRs & LTs exist, but a global vision,
policy & strategy is still missing

Which Communities?

core

- Language Resources
- Language Technologies
- Standardisation
- Content/Ontologies
- System developers
- Integrators
- ...

for

- Many tasks & application domains
 - MT
 - CLIR
 - ...
 - e-government
 - content industry
 - intelligence
 - e-culture

EU
Forum

with

- EC
- National funding agencies
- Industry
- ...

Multilinguality

FLaReNet
Network

Focus on cooperation





FLaReNet at a glance



An international Forum to facilitate interaction, to

- ❑ **Overcome the fragmentation** & recreate a community
- ❑ Create a **shared policy** in the field of LRs and LT for the next years
- ❑ Foster a **European strategy** for consolidating the sector and enhancing competitiveness

Considering that LRs & LTs

- Present **various dimension:** technical, but also organisational, economic, legal, political, ...
- Addresses also **multicultural & multilingual aspects**
 - essential when facing access and use of digital content in today's Europe
- Must be approached from many perspectives & involve **various communities**

FLaReNet promotes **international cooperation**

- E.g. with the US SILT initiative towards sustainable interoperability for LT
 - ❑ E.g. for a **world-wide effort to build consensus about sharing of data and technologies**





LRs, LTs & the Landscape



- **Definition of LRs: very broad**

- Intertwined with LTs & also with usages, tasks, applications
 - Implications on the involved Community/ies
- Definition: today & evolving
 - Wikipedia of LRs

- **The landscape today and in the future**

- **Coverage:** both in the vertical & the horizontal dimensions
 - Portability, Interoperability, Basic Language Resource Kit, Multilinguality, ...
- Which are the big challenges? The realistic ones? Those with more impact?
- Driving forces? External factors? Against & in favour?
- Main objectives?
- ...





Mission



The **largest Network of LR & LT players**

- With diverse approaches

To structure the area of LR & LT of the future by discussing **new strategies to:**

- Enable progress toward **community consensus**
- **Recast the definition of LRs** in the light of recent scientific, methodological, technological, social developments
- **Consolidate** methods and approaches, common practices, frameworks and architectures
- **Integrate** so far partial solutions into broader infrastructures
- Convert existing LTs related to LRs into useful economic & societal benefits
- **Anticipate the needs of new types of LRs / LTs or Infrastructures**





Expected Outcome & Impact



The outcomes will be **of a directive nature**

A “roadmap”

- Identifying areas where consensus is achieved/emerging vs. areas where more discussion & testing is required,
 - Indicating **priorities**
- ➡ Recommendations in the form of a **plan of coherent actions for the EU, national organisations & industry**
- As **input to policy development at EU and national level**
 - Identifying **new language policies supporting linguistic diversity**
 - **Strengthening the language product market, e.g. for new products & innovative services**
- ➡ A (Eu) model for the **LRs/LTs of the next years**

Ambitious!





A Layered Structure

Leading experts & groups (national, European, international institutions, SMEs, large companies) for all relevant LR areas (**about 40 members so far**)

- ◆ in collaboration with CLARIN, EuroMatrix, NSF INTEROP SILT & other world-wide initiatives
- ◆ to **ensure coherence of LR-related efforts in Europe**

Already being enlarged

- ◆ with **new institutional members** &
- ◆ the involvement of the community, as **supporters/subscribers**





First FLaReNet Members



- ILC-CNR - Italy (***Coordinator***)
- ELDA - France
- ILSP - Greece
- LIMSI-CNRS & IMMI - France
- Universitat Pompeu Fabra - Spain
- Universität Wien - Austria
- Universiteit Utrecht - The Netherlands
- Max Planck - Germany
- Bielefeld University - Germany
- Morphologic - Hungary
- Prague University - Czech Republic
- Sheffield University - United Kingdom
- DFKI - Germany
- CEA LIST - France
- Copenhagen University - Denmark
- The Open University - United Kingdom
- Bulgarian Academy of Sciences - Bulgaria
- Vrije Universiteit - The Netherlands
- Romanian Academy - Romania
- Manchester University - UK
- Jožef Stefan Institute - Slovenia
- Adam Mickiewicz University - Poland
- Linguattec - Germany
- Langenscheidt - Germany
- Fondazione Bruno Kessler - Italy
- University of Antwerp - Belgium
- Universitat Politècnica de Catalunya - Spain
- University of Helsinki - Finland
- Universität Tübingen - Germany
- European Parliament - Luxembourg
- DGT - EC DG Translation - Luxembourg
- Vassar College - USA
- Kyoto University - Japan
- Brandeis University - USA
- LDC - Linguistic Data Consortium - USA
- National Institute of Advanced Industrial Science and Technology - Japan
- Korea Advanced Institute of Science and Technology - Korea
- Tsinghua University - China



FLaReNet Steering Committee

- Nicoletta Calzolari, ILC-CNR, Italy (*Coordinator*)
- Khalid Choukri, ELDA, France
- Stelios Piperidis, ILSP, Greece
- Joseph Mariani, LIMSI-CNRS & IMMI, France
- Núria Bel, Universitat Pompeu Fabra, Spain
- Gerhard Budin, Universität Wien, Austria
- Jan Odijk, Universiteit Utrecht, The Netherlands



Universiteit Utrecht





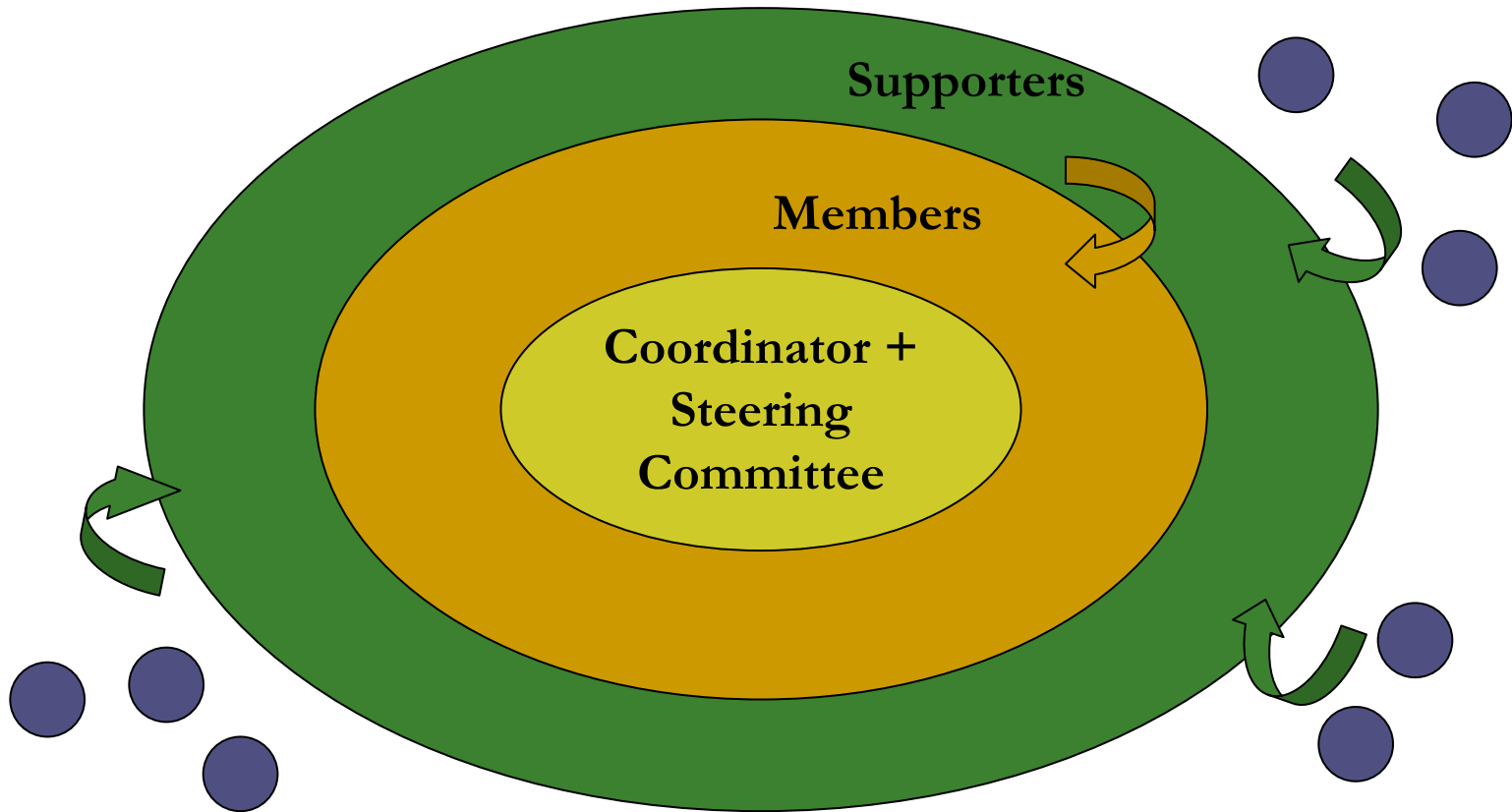
FLaReNet *First* Supporters



- Telefónica Investigación y Desarrollo, Spain
- Neurolingo, Greece
- Archetypon S.A. Information Technologies, Greece
- LIBRIS-TECH S.A., Greece
- Innoetics Ltd, Greece
- Vecsys, France
- The Royal Institute of Technology (KTH), Department of Speech, Music and Hearing, Sweden
- ALPHABET Translation & Interpreting Services EPE, Greece
- ESTeam AB, Sweden
- Mikro Vaerkstedet, Denmark
- QUALIA Technologies of understanding, Greece
- European Bioinformatics Institute (EMBL-EBI), United Kingdom
- Agence France Press (AFP) Medialab R&D Unit, France
- Bertin Technologies, France
- University of Edinburgh, United Kingdom
- Translendum SL, Spain
- ELSNET Foundation, The Netherlands
- DAEDALUS-Data, Decisions and Language, S.A., Spain
- SYNTHEMA S.R.L. – Soluzioni Informatiche, Consulenza e Ricerca, Italy
- France Telecom R&D, France
- European Media Laboratory GmbH, Germany
- Real Academia Española, Spain
- Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento em Lisboa, Portugal
- Norwegian Language Council, Norway
- Siemens AG, Germany
- ...
- ...
- Now many others



The Network is increasing!





Organised in Thematic Areas *Working Groups*



- The **Chart for the area of LRs and LT** in its different dimensions
- Methods and models for **LR building, reuse, interlinking, maintenance, sharing, distribution...**
- **Harmonisation** of formats and standards
- Definition of **evaluation and validation protocols** and procedures
- Methods for the **automatic construction and processing** of LRs



A collaborative
web-site





Work methodology

Involving the community at large



To ensure the **bottom-up** (consensus-based) nature

- survey and analysis phase (preparatory phase)
- discussion phase in **WG meetings** finalized to reaching consensus
- drafting of preliminary recommendations
- external evaluation & feedback, through Support Group, other projects, external experts, ...
- final recommendations
- dissemination of recommendations

FLaReNet will achieve its objectives through

- a **collaborative website &**
- **workshops, meetings**



Meetings are
the F. engine!





FLaReNet Launching Event



The European Language Resources and Technologies Forum:

Shaping the Future of the Multilingual Digital Europe

Vienna, 12-13 February 2009

See the website for the presentations & slides

- To address some technological, market & policy **challenges** to be faced **in a multilingual digital Europe**
 - Occasion for identifying the ground for (& influence) future directions
- Sessions with focused interventions, some organised discussants
 - Large space for discussion
- To elicit *new ideas* from the widest possible audience
 - Reflects the “life-style” of F.
- To play a role in defining recommendations for future actions
 - Shaping the future with respect to the new challenges





International Vocation

- Calls for **international cooperation also outside Europe**

To set up a global

- **Worldwide Forum for Language Resources and Language Technologies**





Results from Vienna: International Cooperation



Standards & Interoperability

- Standards, interoperability & metadata are topics to be approached in cooperation. A metadata catalogue should involve every party
- Common repositories for tools & language data should be established that are universally and easily accessible by everyone
 - Try to connect ongoing work done by many groups
 - The creation of a shared repository with data formats, annotations, etc. – where to find the most frequently used and preferred schemes – is proposed as a major help to achieve and promote standardisation
- Coordinate input to ISO standardisation work also from Asian countries





Results from Vienna: International Cooperation



For a new world-wide language infrastructure

- The issue of access to LRT is a critical one that should involve – and have impact on – all the community
- Need to create the means to plug together different LR& LT, in a **web-based resource and technology grid**
 - With the possibility to easily create new workflows
 - Create conditions to easily share and re-use technologies, to have more open source tools to be made available for use also to under-funded groups.
- A platform for cooperation could be thought of around the notions of BLARK and ELARK





Results from Vienna: International Cooperation



Networking & International Forum

- A focus of FLaReNet and SILT has to be to find a way to really bring everyone on board, making sure all the players around the world are involved
- Networking & support actions must be conducted more intensively, with establishment of international committees that have formal recognition, organisation and participation to common workshops
- International Forum (a meta-body) to share information, discuss strategies and declare that there are common objectives





Actions for FLaReNet



- Use its collaborative website to create a pool of ideas on which to have a joint reflection
- Promote & help in the **standardisation**-oriented tasks and efforts toward harmonisation, sharing and distribution
- Assemble a broad community of relevant people & institutions around the world into a **collaborative network** where institutions & individuals involved are committed
- Promote a **new worldwide language infrastructure** for easy access to LRT, in a web-based resource and technology grid
- Act as a communication vector for **open source** resources and tools: this could be in wiki mode
- Produce a **White paper** summarising ideas for directors of programs of funding agencies, and organise a Forum of directors of funding agencies
- Must establish an **International Advisory Board**: this group can constitute the nucleus & act as the needed International Forum
- Prepare a **MoU** with the main issues discussed and ask members of FLaReNet to sign it when joining the Network





Join FLaReNet!

- We invite all interested players in the field to express their interest in becoming part of the Network

- **How to join?**
 - To be part of the FLaReNet Network fill the form available on the project website

 - (<http://www.flarenet.eu>)

