

The FLaReNet Thematic Network: A Global Forum for Cooperation

Nicoletta Calzolari

Consiglio Nazionale delle Ricerche
Istituto di Linguistica Computazionale
“A. Zampolli”

nicoletta.calzolari@ilc.cnr.it

Claudia Soria

Consiglio Nazionale delle Ricerche
Istituto di Linguistica Computazionale
“A. Zampolli”

claudia.soria@ilc.cnr.it

Abstract

The aim of this short paper is to present the FLaReNet Thematic Network for Language Resources and Language Technologies to the Asian Language Resources Community. Creation of a wide and committed community and of a shared policy in the field of Language Resources is essential in order to foster a substantial advancement of the field. This paper presents the background, overall objectives and methodology of work of the project, as well as a set of preliminary results.

1 Introduction

The field of Language Resources and Technologies has been developing for years to reach now a stable and consolidated status, attaining the right to be considered a discipline in itself, and as testified by the number of conferences and publications explicitly dedicated to the topic. Even if Language Resources (in the widest sense, i.e. spoken, written and multi-modal resources and basic related tools) have a rather short history, they are nowadays recognized as one of the pillars of NLP. The availability of adequate Language Resources for as many languages as possible is a pre-requisite for the development of a truly multilingual Information Society.

At the same time, however, the discipline has seen considerable fragmentation during those years of fast and enthusiast development, and the landscape is now composed by a kaleidoscope of different, often conflicting initiatives that vary as for research directions, theoretical approaches, implementation choices, distribution and access policies, languages, domain and modalities covered, etc.

The growth of the field in the last years should be now complemented by a common reflection and by an effort that identifies synergies and overcomes fragmentation. The consolidation of the area is a pre-condition to enhance competitiveness at EU level and worldwide. There is the need of working together to define common strategies and to identify priorities for the field to advance. Multiple concurring signs are now indicating that time is ripe for establishing an open language infrastructure, something that many of us have been pushing since some time and that is now increasingly recognized in the community at large as a necessary step for building on each other achievements.

Such an open infrastructure can only be realized if the Language Resources community is cohesive enough to be able to focus on a number or priority targets and collectively work towards them, and, at the same time, whether it is powerful enough to permeate the user community, the industry, and the policy-makers.

2 Why FLaReNet

Creation of the necessary conditions for the development of such an infrastructure cannot rely on research activities only and even more cannot rely on the initiative of individual groups. Instead, strategic actions are crucial, such as making contacts with and involving all interested parties, sensitize the policy makers and institutional bodies, involve associations and consortia, disseminating widely the results of common efforts. Only by mobilizing this wide and heterogeneous panorama of actors can such an ambitious goal be attained.

FLaReNet – Fostering Language Resources Network – is a Thematic Network funded by the

European Commission under the eContentPlus framework (ECP-2007-LANG-617001)¹. The FLaReNet Thematic Network was born with the specific aim – as required by the European Commission itself – to enhance European competitiveness in the field of Language Resources and Technologies, especially by consolidating a common vision and fostering a European strategy for the future. A major, long-term objective – as well as a powerful means for community creation – is creating the preparatory environment for making an open language infrastructure a reality.

3 Objectives

The objectives of FLaReNet are threefold:

- The creation and mobilization of a unified and committed community in the field of Language Resources and Technologies;
- The identification of a set of priority themes on which to stimulate action, under the form of a roadmap for Language Resources and Technologies;
- The elaboration of a blueprint of priority areas for actions in the field and a coherent set of recommendations for the policy-makers (funding agencies especially), the business community and the public at large.

3.1 Creation of a community

FLaReNet has the challenging task of creating a network of people around the notion of Language Resources and Technologies. To this end, FLaReNet is bringing together leading experts of research institutions, academies, companies, funding agencies, with the specific purpose of creating consensus around short, medium and long-term strategic objectives. It is of foremost importance that the FLaReNet Network be composed of the as widest as possible representation of experiences, practices, research lines, industrial and political strategies; this in order to derive an overall picture of the field of Language Resources and Technologies that is not limited to the European scenario, but can also be globally inspired. The Network is currently composed of around 200 individuals belonging to academia, research institutes, industries and government. Such a community

also needs to be constantly increased in a concentric way that starts from the core disciplines but gradually projects itself towards “neighboring” ones, such as cognitive science, semantic web, etc.

3.2 Identification of priority themes

Language technologies and language resources are the necessary ingredients for the development of applications that will help bridging language barriers in a global single information space, in a variety of means (the Web as well as communication devices) and for a variety of channels (spoken and written language alike). It is of utmost importance, however, to identify priorities as well as short, medium, and long-term strategic objectives in order to avoid scattered or conflicting efforts.

The major players in the field of Language Resources and Technologies need to consensually work together and indicate a clear direction and priorities for the next years.

3.3 Elaboration of a blueprint of actions

However, whatever action cannot be implemented on a long term without the help of the necessary financial and political framework to sustain them. This is even most true for actions regarding Language Resources that typically imply a sustained effort at national level. To this end, the FLaReNet Network must propose the priority themes under the form of consensual recommendations and a plan of action for EC Member States, other European-wide decision makers, companies, as well as non-EU and International organizations.

FLaReNet goals are very ambitious and its objectives are to be seen in a more global framework. Although they are shaped by the European landscape of the field of LR&T, its mission is therefore inherently cross-boundary: in order to attain such goals getting a global view is fundamental.

To this end, it is important that FLaReNet is known by the Asian community, and it knows the Asian community. Some Asian community players are already members of the Network.

4 How FLaReNet works

Work in FLaReNet is inherently collaborative.

Its means are the following:

- Working groups
- Organization of workshops and meetings

¹ <http://www.flarenet.eu>

- External liaisons

4.1 Working Groups

Working Groups are intended as “think-tanks” of experts (researchers and users) who jointly reflect on selected topics and come up with conclusions and recommendations. The Working Groups are clustered in thematic areas and carry out their activities through workshops, meetings, and via a collaborative Wiki platform. The FLaReNet Thematic Areas are:

- The Chart for the area of Language Resources and Technologies in its different dimensions
- Methods and models for Language Resource building, reuse, interlinking, maintenance, sharing, and distribution
- Harmonization of formats and standards
- Definition of evaluation and validation protocols and procedures
- Methods for the automatic construction and processing of Language Resources.

4.2 Organization of workshops and meetings

Meetings and events lie at the core of FLaReNet action plan and dissemination strategies. They can either be specifically oriented to the dissemination of results and recommendations (*content-pushing* events) or, rather, to their elicitation (*content-pulling* events). Three types of meetings are envisaged:

- Annual Workshops, such as the “European Language Resources and Technologies Forum” held in Vienna, February 2009
- Thematic Workshops related to the work of Working Groups
- Liaison meetings (e.g. those with NSF-SILT, CLARIN, ISO and other projects as the need may arise).

Annual workshops are targeted to gather the broad FLaReNet community together. They are conceived as big events, and they aim at becoming major events in the Language Resources and Technology community of the kind able to attract a considerable audience. Given the success of the formula exploited for

the FLaReNet “Vienna Event”², it is likely that Annual workshops will be organized along the same lines. However, this type of event cannot be repeated on a frequent schedule. At the same time, more focused events centered on specific topics and with extensive time allocated for discussion are essential.

To this end, Annual Workshops will be complemented by many Thematic workshops, i.e. more focused, dedicated meetings with a more restricted audience. These are directly linked to the work being carried out by the various Working Groups and are organized in a de-centralized manner, by direct initiative of the Working Group or Work package Leaders. In an attempt to increase FLaReNet sensitivity to hot issues, selection of topics and issues to be addressed will be also based on a bottom-up approach: FLaReNet members and subscribers are invited to submit topics of interest either freely or as a consequence of “Call for topics” related to particular events.

Finally, liaison meetings are those elicited by FLaReNet to make contact and create synergies with national and international projects that are partially overlapping with FLaReNet in either their objectives or the target audience. Examples of these are the FLaReNet-CLARIN and the FLaReNet-SILT liaison meetings.

4.3 External liaisons

For a Network like FLaReNet, whose aim is the development of strategies and recommendations for the field of Language Resources and Technologies, coordination of actions at a worldwide level is of utmost importance. To this end, FLaReNet is planning to establish contacts and liaisons with national and international associations and consortia, such as LDC, ISO, ALTA, AFNLP, W3C, TEI, COCOSDA, Oriental-COCOSDA. Specific actions of this kind have started already, such as the International Cooperation Round Table that took place in Vienna. The members of the International Cooperation Round Table will form the initial nucleus of the FLaReNet International Advisory Board.

5 First results and recommendations

More than a hundred players worldwide gathered at the latest FLaReNet Vienna Forum, with the

² <http://www.flarenet.eu/?q=Vienna09>, see the Event Program to get an idea of the event structure.

specific purpose of setting up a brainstorming force to make emerge the technological, market and policy challenges to be faced in a multilingual digital Europe.

Over a two-day programme, the participants to the Forum had the opportunity to start assessing the current conditions of the LR&T field and to propose emerging directions of intervention.

Some messages recurred repeatedly across the various sessions, as a sign both of a great convergence around these ideas and also of their relevance in the field. A clear *set of priorities* thus emerged *for fostering the field* of Language Resources and Language Technology.

Language Resource Creation. The effort required to build all needed language resources and common tools should impose on all players *a strong cooperation at the international level* and the community should define how to *enhance current coordination of language resource collection between all involved agencies* and ensure efficiency (e.g. through interoperability).

With data-driven methods dominating the current paradigms, *language resource building, annotation, cataloguing, accessibility, availability* is what the research community is calling for. Major institutional translation services, holding large volumes of useful data, seem to be ready to share their data and FLaReNet could possibly play a facilitating role.

More efforts should be devoted *to solve how to automate the production of the large quantity of resources demanded, and of enough quality to get acceptable results in industrial environments.*

Standards and Interoperability. In the long term, *interoperability will be the cornerstone of a global network of language processing capabilities.* The time and circumstances are ripe to take a broad and forward-looking view in order to establish and implement the standards and technologies necessary to ensure language resource interoperability in the future. This can only be achieved through a *coordinated, community-wide effort that will ensure both comprehensive coverage and widespread acceptance.*

Coordination of Language Technology Evaluation. Looking at the way forward, it clearly appears that *language technology evaluation needs coordination at international level:* in order to ensure the link between technologies and applications, between evaluation campaigns and projects, in order to conduct evaluation campaigns (for ensuring synchrony or for addressing the influence of a

component on a system on the same data), in order to produce language resources from language technology evaluation, or to port an already evaluated language technology to other languages (best practices, tools, metrics, protocols...), in order to avoid “reinventing the wheel”, while being very cautious that there are language and cultural specificities which have to be taken into account (tone languages, oral languages with no writing system, etc).

Availability of Resources, Tools and Information. *Infrastructure building* seems to be one of the main messages for FLaReNet. *For a new worldwide language infrastructure the issue of access to Language Resources and Technologies is a critical one* that should involve – and have impact on – all the community. There is the need to create the means to plug together different Language Resources & Language Technologies, in an *internet-based resource and technology grid*, with the possibility to easily create new workflows. Related to this is *openness and availability of information.* The related issues of *access rights and IPR* also call for cooperation.

6 Join FLaReNet

In order to constantly increase the community of people involved in FLaReNet, as well as to ensure their commitment to the objectives of the Network, a recruiting campaign is always open. People wishing to join the Network can do so by filling an appropriate web form available on the FLaReNet web site. The FLaReNet Network is open to participation by public and private, research and industrial organizations.

7 Conclusions

The field of Language Resources and Technologies needs a strong and coherent international cooperation policy to become more competitive and play a leading role globally. It is crucial to discuss future policies and priorities for the field of Language Resources and Technologies – as in the mission of FLaReNet – not only on the European scene, but also in a worldwide context. Cooperation is an issue that needs to be prepared. FLaReNet may become one of the privileged places where these – and future – initiatives get together to discuss and promote collaboration actions.