FLaReNet Recommendations and Future Prospects

Invited talk

November 25, 2011 at 10:35 (Collegium Iuridicum Novum - Auditorium)

Nicoletta Calzolari

Istituto di Linguistica Computazionale “Antonio Zampolli” – CNR
Pisa, Italy

glottolo@ilc.cnr.it

Abstract

Language Technologies (LT), together with their backbone Language Resources (LR), provide an essential support to the challenge of Multilingualism and Information and Communication Technology of the future. LTs are critical to bridge language barriers and to help creating a new environment where information flows smoothly across frontiers and languages, no matter the country, and the language, of origin.

LT is a data intensive field. To develop LT, research and industrial teams need access to (i) a volume of data commensurate with the operational conditions of the application and (ii) equally appropriate LT evaluation mechanisms to measure the performance of the best systems available and compare it to the performances required by the application.

The LR field is very active, but it needs coherence. Of course, coherence is better assured by sharing common priorities and endeavours. FLaReNet\footnote{FLaReNet is an EC eContentplus project (ECP-2007-LANG-617001)} – Fostering Language Resources Network (http://www.flarenet.eu/) – is an international Forum, composed by a steadily enlarging community, aiming at developing the needed common vision and fostering a European strategy for consolidating the sector, thus enhancing competitiveness at EU level and worldwide. Its goals are:

- facilitate interaction among LR and LT stakeholders and ultimately re-build a community around LRs and LTs, as well as promote and sustain international cooperation
- coordinate a community-wide effort toanalyse the sector of LRs and LTs along all the relevant dimensions: technical and scientific, but also organisational, economic, political, cultural, societal and legal
- identify short, medium, and long-term strategic objectives and provide consensual recommendations in the form of a plan of action targeted to a broad range of stakeholders, from the industrial and scientific community to funding agencies and policy makers
- pave the way to the set up and functioning of an Open Resource Infrastructure through a number of preparatory initiatives.

FLaReNet outcomes are essentially of directive nature: to help the EC, national funding agencies, industrial and academic communities, to identify priority areas of LRs and LTs that need (public) funding to develop or improve.

FLaReNet – by acting as an aggregator and facilitator of information sharing and discussion – has brought together leading experts of research institutions, academies, companies, consortia, associations, funding agencies, public and private bodies both at European and international level. Broad participation of different key-figures ensures that recommendations are formulated through a consensual bottom-up process in which the relevant scientific, technical, organisational, strategic aspects and positions are taken into account.

FLaReNet focused on the field as a whole, its gaps and priorities, in a word a vision for its future. Its final results, described in the final deliverable “Language Resources for the Future – The Future of Language
Resources. The Strategic Language Resource Agenda”, constitute a preliminary plan for actions and infrastructures that could become the basis for future initiatives in the field.

Recognising that the development of the sector of LTs is conditioned by various factors, the FLaReNet recommendations cover a broad range of topics and activities, spanning over production and use of LRs, licensing, maintenance, sustainability and preservation issues, infrastructures for LRs, resource identification and sharing, evaluation and validation, interoperability and policy issues.

In the Blueprint, the various actions recommended are organised around nine dimensions that are relevant for the field of Language Resources: a) Infrastructure, b) Documentation, c) Development, d) Interoperability, e) Coverage, Quality and Adequacy, f) Availability, Sharing and Distribution, g) Sustainability, h) Recognition and i) International cooperation. Some of these dimensions are of a more infrastructural nature, some are more related to research and development, some yet more to political and strategic aspects, but they all must be seriously considered when making up a strategy for the future of the field. All of them eventually have an impact in the development and success of LRs, and represent the areas where actions need to be taken to make the field of Language Resources and Technologies grow.

It is useful to see the various dimensions as a coherent system where each one presupposes the other, so that action at one of the levels requires some other action to be taken at another one. For instance, open availability of data presupposes interoperability (which in turn is boosted by openness); to discover and develop new paradigms, and for data to be usefully exploited, the availability of large quantities of data requires the ability to link the information carried by data; increased data quantity implies a change in their availability towards openness, etc.

Taken together, these directions are intended to contribute to the creation of a sustainable LRT ecosystem.

I will present some of the high-level recommendations collected in the final Blueprint of Actions and Infrastructures that provides the fundamental recommended actions for the development and progress of LRs and LTs, coming from the FLaReNet community and addressing both the community itself and policy makers. Its main goal is thus to serve as an instrument to support stakeholders in planning for and addressing the urgencies of the LRs and LTs of the future.

A larger and larger range of LRs and LTs is being developed, but the infrastructure that puts LRs and LTs together and sustains them is still largely missing. Infrastructure building is the most urgent issue – as indicated by many – and also a way to make the field move forward, together with real sharing of resources, and an effort to make resources available for all languages. In this context, interoperability of resources, tools, and frameworks has recently come to be understood as perhaps the most pressing current need for language processing research.

Technical scientific issues are obviously important, but organisational, coordination, political issues play a major role in our field as in every other. Technologies exist and develop fast, but the infrastructure that puts them together and sustains them must be properly designed, created and sustained.

The production process of the FLaReNet Blueprint is the result of a permanent and cyclical consultation that FLaReNet has initiated inside the community it represents – with almost about 400 members – and outside it, through connections with neighbouring projects, associations, initiatives, funding agencies and governmental institutions. As a result of this process, the final version of the Blueprint reflects a consensus on the active living interests, needs, concerns, and values of a wide spectrum of players in LTs.

Since its very beginning, FLaReNet has recognised the need of overcoming the European dimension for a truly incisive footprint to be left. A global dimension is to be taken into consideration when planning and designing the future of Language Resources and Technologies.

---