

Putting Language Resources Infrastructure to the Test: the ESFRI Challenge

Tamás Váradi

Linguistics Institute, Hungarian Academy of Sciences
varadi@nytud.hu

The Language Resources and Language Technology community is one of the most dynamically growing vibrant communities of recent years. This is well attested by the history of LREC itself. It did not take ten years for it to become a massive event drawing several hundred contributors. Language resources are clearly seen as a cornerstone of research activities that provide impetus to a number of related fields ranging from hard core ICT projects to general-interest language preservation and querying. Still, it is fair to say that the main driving force behind language resources has been the language technology industry ever craving for more and more data.

The European Strategy Forum on Research Infrastructures (ESFRI) was launched in 2002 with the aim of working out a common platform on research infrastructures in Europe, and “to act as an incubator for international negotiations about concrete initiatives”. The current ESFRI activities are focussed on creating a Roadmap of new research infrastructures of pan-European interest, which is due to be published by the autumn of 2006.

ESFRI members are one of two persons delegated by each member country who are typically high ranking officials in charge of cultural/scientific policy. Preparatory work has been going on in three Steering Groups devoted to Physical Sciences and Engineering, Biological and Medical Sciences, and Social Sciences and Humanities respectively. Their work was helped by Expert Groups consisting of 8 – 10 members. The Social Sciences and Humanities Working Group has two Expert Groups, one to cover Social Sciences the other devoted to Cultural Heritage. ESFRI started out reviewing existing Roadmaps with a view to integrating them in the ESFRI Roadmap but the main source of information for identifying potential infrastructure initiatives for the Roadmap was a questionnaire circulated through ESFRI members, inviting applications for projects to be identified for inclusion in the Roadmap.

There were three major initiatives from the domain of language resources and language technology submitted independently of each other, which in the end were consolidated into a single proposal named CLARIN. The CLARIN proposal

preserved much of the broad community forming objectives of the EARL initiative but focussed its aims on serving the ESFRI SSH community.

This is challenge number one. The ESFRI call is to propose research infrastructure for the social sciences and humanities. Linguistics is clearly within this domain but important as it is, language resources and language technology has a much broader relevance than serving the needs of linguists. In fact, we need to drive this point home with most people outside our field because otherwise the popular view that language resources/technology is about linguistics prevails. But as far as the current ESFRI proposal is concerned, it is not enough to get our aims and scope of relevance clearly established. We need to constantly remind ourselves that CLARIN is not an infrastructure for our own community. It is an infrastructure meant to serve the needs of the social science and humanities researchers. This is a completely new role in that the traditional user base of our community was keen and, most importantly, able to make use of what our field had to offer. In contrast, we can expect no such readiness with humanities scholars. They may not even be aware of the benefits of using language resources and the relevant language technology in their own research. Hence the challenge to build an infrastructure that provide services that are capable of not just making resources and technology available but readily usable as well by the target audience. It requires actively promoting the use of language resources and technology and also providing them in as much tailored to the perceived needs of the target audience as possible. I am sure the discussion in the present workshop will result in many useful ideas for the strategies to follow to achieve this objective.

The other challenge concerns financial viability. The ESFRI Roadmap is supposed to be a select choice of initiatives that have the seal of approval by ESFRI as meeting the criteria of the scientific soundness and viability, pan-European relevance and maturity. Being on the Roadmap means no guarantee for being funded by the EU. In fact, ESFRI depends on the willingness of member countries to buy into the projects identified on the Roadmap. The biggest challenge for CLARIN, then, is to round up enough national support to make the project economically

feasible with minimal EU contribution. This is, again, an issue that calls for detailed discussion in our workshop.