

## **A EUROPEAN LANGUAGE RESOURCES INFRASTRUCTURE**

**Josep Soler**

**European Commission**

**Directorate General XIII Telecommunications,  
information market and exploitation of research**

The media characterise the 90's by the exponential increase of information and communication systems that have changed the way we live and work. As a response to this challenge, we have already seen the development of new information services capable of managing the flow of information that is the hallmark of our society, making it easier for each and every citizen to access this information. As the economic indicators remind us, we are steadily moving from an industrial towards an information and communication society which evolves in parallel with the expansion of commercial exchanges world wide.

Much more is expected to come and Europe must successfully grow into this new information society if it wants to keep competitive in the world market. To be competitive Europe must care not only for the present but, more importantly, anticipate the future.

As for the present, the Commission's White Paper on "Growth, Competitiveness, Employment" has already stressed the potential impact of information and communication products and services on employment and quality of life, and the so-called "Bangemann Report" has provided a strategic action plan designed to facilitate and accelerate Europe's transition to the global information society, based on the concentration of efforts on key telematic application areas such as Transport, Health Care, Libraries, Language Learning or Language Engineering.

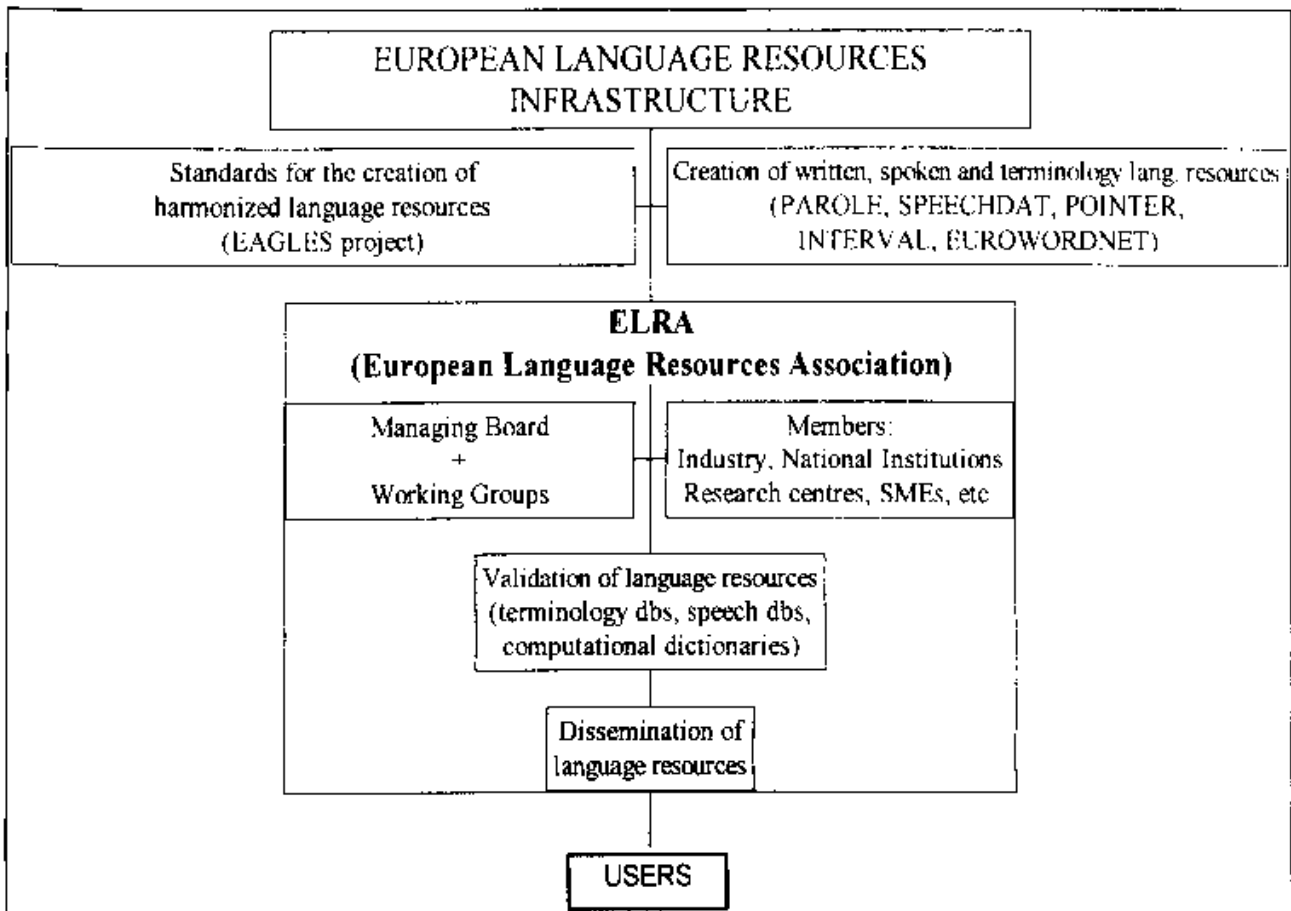
The importance of the Language Engineering sector in the future of the information society, in particular, lies in the fact that a language engineering component is likely to be required by any telematic application to access, for example, a multilingual information system, develop interactive transactions such as telebanking and teleshopping, draft or translate documents via networks, classify information for retrieval purposes, develop voice recognition systems, generate a text automatically, prepare advanced language learning material, create more powerful grammar and spell checkers, or navigate in the WEB by means of the user's mother tongue. The opportunities are largely open.

Transforming language technologies and components into marketable systems, products and services depends not only on the availability of a solid scientific and technology base but also on the availability of large and complete resources such as electronic dictionaries, terminology banks, speech databases, and computational grammars and tools, the essential building blocks on which to develop practical language engineering systems.

In response to a request made by the European Council in Corfu (June 1994), the Commission has been asked to develop the linguistic aspects of its information society proposal in order to promote effective technology-mediated business that contribute to better usability and wider deployment of multilingual telematic products and services. A multilingual approach to communications will provide equal opportunities for European citizens and businesses, whatever their language, to participate in the new information age and to take full advantage of it and will also promote and ensure language diversity. At this respect, each Member State and region will have to play its role in preserving its language or languages, in accordance with subsidiarity.

A programme proposal on those issues is now being prepared by the European Commission under the title "The Multilingual Information Society". It will support the construction of an infrastructure for European language resources, mobilise and expand the language industries, and promote the use of advanced language tools in the European public sector.

In the language resources field, the Commission tries to cope with the present of language engineering through planning and promoting infrastructural measures. The current Language Engineering sector, part of the Telematics Applications Programme of the European



Commission (1994-1998), has already funded some projects aimed at preparing the ground for the large scale development of language resources. The European Commission is also preparing the future of this sector supporting the efforts of various industrial and public bodies aimed at building large-scale, mono- and multilingual language resources – which are widely acknowledged to be critical for the development of robust and cost-effective language applications – for the European languages, in particular for minority ones.

Projects such as PAROLE, EUROWORDNET, POINTER, INTERVAL, and SPEECHDAT are developing textual corpora, lexicons, and speech databases for a large variety of European languages, as well as validation tools for terminology. Besides the considerable technical and scientific results achieved so far, what is equally important, maybe even more important, is that these projects have contributed to design a whole scenario for the development of language resources, that includes standardisation measures for their (EAGLES project) implementation projects (various language engineering

demonstration projects), and mechanisms for the validation and distribution of such resources (this will become possible through ELRA).

ELRA (European Language Resources Association), in particular, will provide an opportunity to promote the emergence of a language resources exchange framework, where suppliers and users, national organisations and research institutions are encouraged to co-ordinate their resources and results in order to ensure that Europe's language industry can benefit to the full.

Setting out an infrastructure for language resources requires adequate time frames and important investments that go well beyond the possibilities of private sector actors, particularly as most of the firms involved in language engineering developments are SME's. The various industries and other organisations that commit themselves to the creation of resources should do so for a period that allows to come to exploitable results. But setting out a linguistic infrastructure also requires good co-operation with the initiatives of Member States, because the development and continuous update of tools and resources for a particular language is primarily the responsibility of the Member States where that language is spoken.

The lack of transnational structures for the creation and dissemination of language resources has undoubtedly been one of the greatest handicaps which Europe has had in providing a basis for the exchange of knowledge and linguistic data, allowing Europeans to participate actively in the multilingual information society. The fragmentation into European and national projects, which has been the rule up to now, has prevented an economically rational division of effort that would stimulate productivity and benefit from synergy.

Co-operation is essential if European competitiveness is to be increased. The European Union is made up of a great variety of different cultures, languages, and traditions. The approach the Commission has worked out in the language resources area seeks to exploit linguistic diversity as a great cultural resource and an important repository of the European heritage.

One could say that we are moving towards the electronic control of information. In view of this, and in view of the fact that language is central in the human communication process, it is crucial that language takes a front seat in the development of our information society.

Human communication in information society must be aided by exploiting technological advances in information technology. Such a move has the potential of improving human communication and the use of language – not just one particular language, but all languages – in a multilingual and multicultural Europe.