

## **Message from ELRA Secretary General and ELDA Managing Director Khalid Choukri**

Welcome to Malta and LREC'2010; Merhba lil Malta u LREC'2010;

Welcome to the celebration of ELRA 15th anniversary; your presence here is the best gift that friends can bring for such a wonderful event; thank you all for your contribution to make this happen;

But let me first express, on behalf of the ELDA team, our profound gratitude to His Excellency Dr George Abela, President of Malta, for his Distinguished Patronage of LREC2010 and for honoring us with his presence. Our gratitude is also extended to His Excellency M. Herman Van Rompuy, President of the European Council for the importance that the EC confer to multilingualism, one of the cornerstones of our conference.

Patronage of His Excellency Dr George Abela, President of Malta highlights the importance his Excellence attributes to this field and the encouragements of his Excellence to the diversity of our topics and I know that this is also a true commitment to see the languages reinforce the foundations of intercultural comprehensions and diversities in Europe and beyond. So in the name of ELRA and of the LREC Program Committee, and I am sure I can speak on their behalf, let me express to his Excellence our gratitude for his support and extend this gratitude to his Excellency, president Van Rompuy, assuring him with our dedication to help with the huge challenge of Multilingualism on his agenda.

I am particularly proud to welcome you in Malta where many of our visions are shared by the country highest authorities but also where our main concerns are explicitly imprinted every where. Maltese (Malti) is the only Semitic language that is an official language of the European Union, it has all the roots and grammar paradigms of a Semitic language while it borrowed from Italian, Sicilian, and more recently from English a large amount of its vocabulary and uses Latin alphabet for its writing. The country population lives the multilingualism at a very high degree, mostly bilingual (Maltese, English), the population has a good command of Italian, a nice scenario to mimic at other corners of the European Union and beyond.

We are happy that this 7<sup>th</sup> LREC is taking place in that context; ELRA and its partners are very proud to continue the organisation of such a major event where one can meet all active players from both academia and industry. Going back to the 1998 vision initiated by Joseph Mariani and Antonio Zampolli for the need of a scientific event dedicated to LRs and Evaluation, I remember that we welcomed less than 500 participants in Granada (1998) followed by Athens (2000), Las Palmas (2002), Lisbon (2004), Genoa (2006) and Marrakech (2008) with increasing interest from the HLT community. We, the organizing team, felt that we reached the highest target in terms of number of participants, papers and submissions, workshops but we have broken a new record with more than 1100 registered participants a month before the Conference, about 9 tutorials and 24 workshops.

This stress how dynamic our field is but also how vital it is in particular for a continent like Europe attempting to establish a Multilingual Union that fundamentally requires HLT to help overcoming the language barriers and bringing a crucial supporting tool for the Challenge of Multilingualism. ELRA has committed to such vision from that start and will continue being supportive to all activities and initiatives that help making LRs available to the R&D communities and also providing the necessary resources and the infrastructures for the evaluation of technologies.

As you know ELRA is driven by its members and enjoyed over the last 15 years the membership of more than 70 active players in average, each year. Gathering such members, that enjoy all ELRA membership benefits, helped also ELRA to design and tune its strategy and orientation. I am particularly happy to welcome the 150 participants to this edition and coming from institutions, members of ELRA.

To offer useful services to its members meant for ELRA to devote human resources to the identification of useful LRs and making them available through its LRs catalogue. From a “centralised” archiving house, as was initially established, ELRA shifted its operations towards brokerage of LRs with strong focus on distribution and licensing. This required, in addition to the negotiation of distribution rights with providers, the drafting of reliable and understandable licences and licensing schemas, ensuring that use (and re-use) of LRs is not hindered by legal barriers. With the support of legal experts, the ELRA team acquired the necessary legal knowledge to handle the basic legal issues involved.

Over the 15 years of activities, more than 1000 resources have been catalogued and made available, thanks to over 250 distribution agreements (with providers from all over the world); ELRA distributed over 3500 resources for HLT development (respectively 48% for research purposes by academia, 37% for Research and technology development by industry, and 16% for evaluation of technologies), not to mention an additional 1500 copies distributed within evaluation campaigns. The availability of such resources within easy, trustable, and fair legal and exchange frameworks boosted the development of Human Language Technologies and the deployment of applications.

In addition to the over 1000 items available off the shelf from ELRA catalogue, ELRA has also continued its work on identifying existing LRs and collected useful information on over 1700 resources that constitute the ELRA Universal Catalogue, an antechamber to the ELRA catalogue and the ancestor of the LREC Map.

A number of HLT evaluation campaigns have been initiated, others have been strongly supported ensuring that the community get a clear picture of the state of the art but also ensuring that all used resources are assembled afterwards and offered to the community as “Evaluation Packages”, with all required datasets, tools, methodologies, to conduct similar experiments. More than 20 technologies have been evaluated with that contexts and over 40 Packages are available within the ELRA Catalogue. In order to ensure that the community has access to all available information on Evaluation initiatives, ELRA has set up a portal collecting and compiling all sorts of information related to evaluation with quick and easy reference about protocols, metrics, tasks, resources, projects, campaigns, etc. (<http://www.hlt-evaluation.org>).

ELRA has also conducted the production of LRs, either on its internal funds or commissioned by partners. Many of these productions took place in the framework of European and international projects. So far ELRA compiled LRs in more than 25 languages, with processes that ensure high quality. Such LRs target different technologies. Some of the recent achievements covered audio/speech data for a variety of languages (e.g. Hindi, Korean, Colloquial Arabic(s), Canadian French, US Spanish, etc.), Broadcast News Speech Corpus for Arabic, French, Spanish, etc. Corpora for languages such as Turkish, Romanian, Kazak, Catalan, etc. aligned textual corpora for Machine Translation (several languages), video annotations with audio transcriptions, etc. etc.

Most of these resources are intended to be part of the ELRA Catalogue, even if, when commissioned, all the rights including ownership remain with the commissioning party.

Last but not least, ELRA continues its work on providing useful forums for the HLT community to debate on current situations, new trends, and shared visions for the future. In addition to LREC, established as a major milestone within our community, ELRA continues to support the organization of the Langtech events (the European exhibition on language technologies, previous one in Rome 2008, next one very likely in 2011), the organization of what is now known as the “European Language Resources and Technologies Forum” (held in the framework of the EC funded project

Flarenet in Vienna, 2009 and Barcelona, 2010), the organization of specialized events (MEDAR conferences on the Arabic language, workshops on Less-Resourced Languages (last one in Poznan, in conjunction with LTC'09), etc. Such events allow the community to gather and discuss the hot topics and ensure a useful stream of information dissemination and awareness.

Now that many of the original core goals have been achieved, others expanded, and many activities are running smoothly, ELRA is moving to a new stage. After an establishment and consolidation period, ELRA is now well equipped to tackle the new challenges emerging from the new trends within our community.

ELRA is reshaping its visions and missions throughout a number of infrastructural changes that will offer our community new services centred on LRs and Evaluation, taking into account new scenarios of LRs "consumption behaviours", while keeping in mind the rationales behind its establishment.

The core activity centred on identification of Language Resources, negotiation of distribution rights and cataloguing such LRs on our online catalogue will remain as our backbone. The modus operandi will strongly take into consideration new mechanisms of sharing and distributing LRs, combining all Web 2.0 features and inspiring both from "open source" principles and professional Business-to-Business approaches.

Part of this vision will be driven by the strong implication of ELRA, through ELDA, in some major European and international initiatives. An international initiative is ongoing to secure the largest Global catalogue of LRs, partnering with LDC, NICT, and others. Others will be driven by a number of EC newly funded projects that have just started e.g. META-NET and PANACEA to name the crucial ones; META-NET, a network of excellence that will design "an open, integrated, secured and interoperable exchange facility for language data and tools for the Human Language Technologies domain" called META-SHARE. Our expectations within META-SHARE is to make more resources sharable through open and integrated Repositories that share similar principles of community servicing and which would benefit from ELRA 15 years of expertise. The other important project, PANACEA, is addressing issues related to cost-effectiveness of the production of LRs and its automation through adequate web-service based "factories".

Another reason to offer a forum for discussion like LREC is to ensure that needs, trends, expectations, plans, partnerships, co-operations can be discussed and brought one step forward. Given the number of initiatives, projects, committees, that address some of the issues of paramount importance to our community, it is becoming essential to harmonize and coordinate the activities of international organizations. I know that many of us share this view and would like to take this opportunity to issue a public invitation to coordinate our efforts, through an "informal" umbrella organization, to all interested organizations (ACL, COLING, IAMT, ISCA, EAFT, LDC, GSK, AFNLP, COCOSDA/ORIENTAL-COCOSDA, etc.). LREC'2010 could be the right place to start such initiative.

If you would like to learn more about ELRA and ELDA, the ELRA/ELDA staff is available during the conference. You will also find more information on our web sites, at [www.elra.info](http://www.elra.info) and [www.elda.org](http://www.elda.org).

On behalf of ELRA and ELDA, and on your behalf, I would like to warmly thank the local team in Malta as well as the Maltese advisory committee.

ELRA and the Organizing Committee have made every effort to ensure the success of LREC 2010, making, both the scientific part and the social events fruitful and enjoyable. Suggestions to improve any aspects of the conference are welcome, and if you need any assistance to make of this event a more memorable one, please do not hesitate to contact our staff, who will be very pleased to help.

Once again, welcome to Malta, welcome to LREC 2010.

Khalid Choukri  
ELRA Secretary General and ELDA Managing Director