Message from ELRA Secretary General and ELDA Managing Director
Khalid Choukri

Welcome to Istanbul and LREC 2012,

Welcome to this LREC 2012, the 8th edition of one of the major events in language sciences and technologies and the most visible service of ELRA to the community.

I would like to extend our warm welcome to the 140 representatives of ELRA members, attending LREC2012.

On behalf of ELRA members and LREC participants, I would like express our gratitude to Ms Neelie Kroes, Vice-President of the European Commission, in charge of the Digital agenda, for her Distinguished Patronage of LREC 2012.

Organizing LREC 2012 under the auspices of these distinguish patrons is an important sign, for us who manage signs, symbols and semantics, regarding the importance conferred to languages, multilingualism, information technologies and all related fields.

These issues are at the heart of EU Digital Agenda, an Agenda that should consider Language Technologies as an essential path to pave the way to automating not only human-machine interactions, human access to information but also human-human communications, across languages and across cultures.

After having organized LREC in Marrakech and Malta, two representatives of Semitic languages (Arabic, Maltese), we are this time in a city that played one of the most noticeable roles in forging Europe, parts of Asia and Africa history and geopolitics, as well as languages, with its own language family, the Turkic languages. After a number of centuries, during which Turkish shared many aspects with Arabic and Persian including a writing system, the foundation of the republic of Turkey came with the script reform (shifting from "Arabic" characters to Latin ones) and the foundation of the Turkish Language Association in 1932 under the patronage of Mustapha Kemal Ataturk himself. The association revived so many Turkic terms and came out with so many neologisms to establish the modern language. This experience, event if not unique in mankind history, is an important process for us, Language scientists and engineers.

At ELRA, we are very happy to carry out and support activities that help all languages to have access to resources essential for their move forward for a bright future and in particular for ensuring access to the digital world and reducing the digital divide.

We are very proud to organize this 8th LREC in that context and for that purpose: to offer our Community the forum it needs, where all players can meet and discuss hot issues related to language resources, technology evaluation, and language sciences.
With more than a thousand participants attending each LREC since 2008, we feel confident that such event where players from Academia and Industry can meet, where new comers, students and junior researchers can find background knowledge and where researchers can review new theories and trends.
With more than 1100 registered participants, more than 30 specialized workshops, about 10 tutorials, almost 700 papers at the main conference, we feel that the achievement is worth the effort we dedicate to make it happen. Boosted by this vitality and energy of our field, ELRA is moving forward with new

objectives and new services to anticipate the community expectations in its challenging task to bring in more supporting tools and automations, to overcome the language and cultural barriers, and help humans enjoy the multilingualism, multiculturalism of the global world of today and tomorrow.

Over the last 17th years (1995-2012), ELRA, driven by its members' instructions, requirements, expectations, has established a number of activities to serve them. LREC is "only" and (probably) the (most) visible aspect of such services.

As many of you know, the core activity of ELRA has been and continues to be identification of valuable Language Resources, useful for research, development and evaluation of Language Technologies. Such identification, followed by a time consuming process of negotiating distribution conditions and clearing all legal issues, led to the constitution of the ELRA catalogue of over 1000 language resources and evaluation packages

In order to help enrich such catalogue, ELRA initiated an identification process to collect and compile data on all existing resources, worldwide, to ensure that such information is shared within the community. This is our Universal Catalogue (UC). UC comprises all identified resources and a priority list is drawn before to launch the negotiations with right holders on sharing and distributing them.

To supplement this, another initiative was launched by ELRA at LREC'2010, the LRE'Map (a Language Resources and Evaluation map). LRE'Map allows each LREC author to describe resources used in his/her work. More than 1200 LR descriptions have been collected at this LREC. LRE'Map feature is now exploited by other conferences and we hope it will become a common feature to all Language Technology events (www.resourcesbook.eu). Such map contributes to spreading and sharing knowledge about LRs.

It is clear that such repositories and resources, along with fair, easy to use, and trustable legal conditions played a role in deployment of Languages Technology applications.

Since 2010, as partially reported on at LREC 2010, ELRA, through its operational body ELDA, is taking part to the META-NET Network of Excellence (Technologies for the Multilingual European Information Society). The main objective is to move forward and extend existing distribution and sharing mechanisms within a new paradigm. For this purpose, the consortium focuses on "Building an Open Resource Infrastructure", for sharing language resources and tools, referred to as META-SHARE.
META-SHARE aims to be "*a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources.*" (cf. http://www.meta-net.eu/meta-share).
One of the essential tasks of the project is related to the metadata issues with respect to the description of LRs. Work has being carried out for the specification of a metadata schema which builds upon available schemas e.g. ELRA, knowledge and expertise and provides a unified schema capable of handling the requirements of the community. These requirements comprise both the description of Language Resources and that of tools or technologies. A large number of Language Technology organizations have been debating the harmonization of such descriptions. In addition, this work aims to consider new modalities such as video and image (for e.g. sign languages, multi-sensor or multi-modal data, etc.). This work on metadata is now mature enough to be considered for standardization. More than 50 players have adopted it and many tools (metadata editor, converters from existing schemas, etc.) are made widely available.

A related issue on which ELRA and a large number of Language Technologies organization have been debating is the harmonization of the identification of LRs. A consensus seems to emerge regarding the set-up of a small executive committee, steered by a commission representing all key players in the field, data centers (ELRA, LDC, Allagin,/GSK, C-LDC,…), and the stack holders (ACL, IAMT, ISCA,…), to assign each LR an International Standard Language Resource Number (ISLRN), independently of whether the LR is accessible on Internet, Intranet, available or not, etc… whether it

has a DOI, a local PId, etc. Such ISLRN should guarantee that all LR usable within our field get a unique identifier that can be used to distinguish it from others.

Another important aspect, the harmonization of existing licensing schemas and the legal aspects, has been part of the discussions and in particular, ELDA focused on the commonalities between ELRA licenses and the ones promoted by Creative Commons, with the intention to harmonize such licenses under the new umbrella of META-SHARE, which was done and will be debated during this LREC at a dedicated workshop.

A version of the META-SHARE network of repositories is already available (www.meta-share.eu) and more information about it is provided in the ELRA's president message as well as at the corresponding LREC workshop, tutorial, and several accepted papers.

As indicated above, a major barrier that hinders the sharing of language resources and tools is the copyright and other IPR issues. ELRA and the META-SHARE partners have been working hard to offer a harmonized set of licenses that cover all needs and sharing/distributing scenarios. In parallel, ELRA continues to advocate for simplifying copyright and IPR issues concerning LR, in particular when used for research purposes. Such exception, which exists in a number of countries (e.g. section 107 of the US copyright law), deserves to be harmonized and extended to all countries. LREC offers a useful forum for debating such issue and hopefully coming up with a common declaration on this and other similar hot topics, to be pushed forward by all of us back home.

This has been a strong credo of the FLaReNet project (in which ELRA Board members and many stack holders including ELDA) took an active role. FLaReNet conclusions at its annual forums, advocated for this harmonization. It went beyond that and compiled a useful but critical roadmap, available to all (www.flarenet.eu), and drew a clear picture of the new trends and important expectations and paved the path for ELRA activities for the coming years. Its recommendation on "Language Resources for the Future – The Future of Language Resources, The Strategic Language Resource Agenda" is an essential roadmap for us.

One of the conclusions that has been thoroughly debated within the board of ELRA is the set-up of a new permanent forum, gathering all LREC attendees and all interested individuals to constitute the Language Resources and Evaluation Forum (LRE-F). We feel that it is important to identify and gather the members of this very broad community and ensure that interactive exchanges/services can be set up to help them work together. The forum is established at this LREC 2012 where the largest group of individuals that have to do with Language Resources and Evaluation are present; it is open (and not limited) to: scientists, students or professors, involved in research activities in universities, small and medium companies or international groups; decision-makers or project managers in large public institutions, etc. You have been invited to join when registering for LREC and we hope you expressed your wish to join. Those who missed that opportunity still can do so at any time through the ELRA portal. Among the services, members of the LRE-F will be offered free downloading of many resources from the ELRA Catalogue and the META-SHARE repository, access to the legal helpdesk, access to the LRE Map, the LR Library, access to LRE Wiki, etc. Members of the community will be also encouraged to join so to upload resources on the ELRA and/or ELRA-META-SHARE repository to share with other colleagues.

An additional service offered by ELRA to all its partners, is the production, customization, repurposing of Language Resources, on demand. ELRA, through the ELDA staff, is involved in LR production. Such productions comprised speech corpora, lexica, textual corpora, both monolingual and aligned / comparable multilingual ones, video and audio data, documents and many other modalities. Such activities included production from scratch as well as, repurposing of existing ones, merging of various sets, annotations, transcriptions, META-DATA labeling of existing databases, etc. ELRA carried out such production for more than 30 languages, working proudly with hundreds of local partners all over the world.

In order to turn this into efficient and cost-effective services, ELDA is part of the EC project PANACEA (Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources). The project aimed at building a factory of Language Resources that "*progressively automates the stages involved in the acquisition, production, updating and maintenance of language resources*", in particular those required by MT systems. The platform will be available both as a Framework (software package) for partners to deploy and as a service offered by ELRA for specific production of resources.

ELRA is ready to assist in LR productions, at any of the needed stages.

ELRA continues to produce resources for technology evaluation and the related campaigns. We would like to stress the importance of packaging the LR and methodologies used for such purposes, to help other interested colleagues in carrying similar assessments. It is also crucial to review such resources for possible repurposing for other needs. ELRA is prepared to assist all evaluators in these tasks. More than 50 packages are already available through ELRA catalogue, most of them for free. In order to keep an efficient stream of information on this, ELRA continues to support the HLT evaluation portal (www.hlt-evaluation.org).

While preparing this message, I went back to messages of our first gathering in Granada (LREC'1998), ages ago one would think!

> *"The presently embryonic infrastructure should be reinforced, so that the same infrastructure is able to coordinate and perform, avoiding duplications, different complementary tasks: to provide and update the general repertories of linguistic data and knowledge which should be available for as many languages as possible, to produce at reasonable costs and in due time customized LR to answer specific requests of developers, to offer services the community urgently needs, information, consultation, validation, etc. "* (Antonio Zampolli, Introductory message to LREC 1998, Granada)

After the set-up and consolidation of ELRA, and now with our strong commitment to boost and sustain META-SHARE, we feel these new approaches to efficient and cost effective sharing of LRs are essential milestones for our community and ELRA is very proud to play a role in this effort.

Last but not least, let me tell you a few words about our week here in Istanbul. In addition to the technical and scientific program (see more details in our LREC Chair message, herein), we have designed, with our local colleagues, a social program to make our stay enjoyable but also fruitful for establishing new relationships and networks, setting up new projects and collaborations, and above all making new friends.

We did our best to make your stay in Istanbul a very pleasant experience, we hope that both our welcome reception (Wednesday, May 23) and Gala Dinner (Friday, 25 may) will give you memories to treasure. We hope that during these events and throughout the week, we will show you some of the best Turkey as to offer.

As always, we tried to introduce novelties and new features to improve the organization of LREC.

In addition to the EU Village, a dissemination / exhibition opportunity for EU projects, we have extended this with an EU "track" of oral presentations, to offer you a full afternoon of information on the major activities supported by the EC (Thursday, May 24).

LREC 2012 will definitely close the chapter of proceedings supplied as hardcopies, CDs or USBs. We will keep the tradition to provide the participants with hardcopies of the program booklet, and the abstracts (of papers of the main conference and the workshops, material of tutorials). BUT the proceedings will only be made available and in advance, on the LREC web site, and in various format, so that you can download them on your favorite media and bring them with you. Please do that in advance, local Internet connection may not be efficient enough for all of us to do that locally.

A new experiment will be conducted this time, a tool, called MyLREC-program, will allow participants to choose their sessions (even the papers they would like to hear within a given session), design their own program and plan their days. One can print it as a PDF file or import it in one's favorite calendar. Please visit the LREC2012 pages for this. We hope this will help you navigate efficiently and friendly through all the sessions LREC is offering.

Finally, I wish to express my deep thanks to our partners and supporters, who throughout the years make LREC so successful.

I would like first to thank our Silver Sponsors: CELI, NUANCE; our bronze sponsors: EML (The European Media Laboratory GmbH), IMMI, K-dictionaries, META-NET, and Quaero.

I would like also to thank the EC Village participants; we hope that such gathering will offer them an opportunity to foster their dissemination and hopefully discuss exploitation plans with the attendees.

I would like to thank the LREC Local Committee, chaired by Mehmed Özkan, who helped us with all logistic issues.

I would like finally to warmly thank the joint team of the two institutions that devote so much effort over months and often behind curtains to make this one week memorable: ILC-CNR in Pisa and my own team, ELDA, in Paris. These are the two LREC coordinators Sara Goggi and Hélène Mazo and the team: Victoria Arranz, Cécile Barbier, Paola Baroni, Roberto Bartolini, Riccardo Del Gratta, Francesca Frontini, Olivier Hamon, Valerie Mapelli, Vincenzo Parrinelli, Valeria Quochi, Caroline Rannaud, Irene Russo, Priscille Schneller.

LREC is yours; we hope that each of you will achieve valuable results and accomplishments. We, ELRA and ILC-NCR staff, are at your disposal to help you get the best out of it.

Once again, welcome to Istanbul, welcome to LREC' 2012

Khalid Choukri
ELRA Secretary General and ELDA Managing Director