

Message from ELRA CEO

Khalid Choukri

Dear LREC participants,

Welcome to LREC 2008, welcome to Marrakech,

ELDA (Evaluations and Language resources Distribution Agency), the operational body and distribution agency of ELRA, and its Moroccan affiliated partner MLTC (Med Language Technology Center, Rabat, Morocco) are very highly honoured by His Majesty the King Mohammed VI, King of Morocco, who kindly agreed to Patronise LREC 2008. On your behalf we express our profound gratitude to His Majesty for His Royal Patronage that highlights the importance His Majesty assigns to multilingual and multicultural issues.

We are proud to welcome you in Marrakech, where we are pleased to contribute to the organization of this sixth edition of the Language Resources and Evaluation Conference, LREC 2008. We are very happy to continue the organization of such a major conference for the Human Language Technology community in such a wonderful city.

Some of you may remember how things started ten years ago in Granada (LREC 1998). At that time we were in the twentieth century but, for our community and field, it felt almost like the middle ages, not only in what regards the progress made in language technologies but essentially regarding the availability of language resources, the initiatives concerned with the sharing of resources, and the evaluation campaigns we see in many areas and all over the world today.

Granada and LREC started after a meeting held in Madrid by a small group of visionary people around Antonio Zampolli and Angel Martin-Municio, to whom we pay tribute at every LREC. Both of them realised about the huge gap that existed in our field (referred to as Language Engineering at the time, strongly influenced by the European Commission jargon on those days). That small "task force" started thinking of an event that would gather the community and that would focus on Language Resources and Language Technology evaluations, the core activities of ELRA. This first edition attracted more than 500 participants and about 200 papers were accepted. None of us expected such great interest from the start but it proved the need for such a forum and strengthened our willingness to continue. With the number of papers steadily increasing, we had to think of different ways to organize LREC. Most of the work is now shared between Pisa (ILC), Paris (ELDA) and our local partners. The organizers' best reward is this increasing number of papers and participants. At LREC 2006 in Genoa we had more than 500 papers and over 875 attendees. For LREC 2008 in Marrakech we have received more than 900 submissions for about 650 accepted communications and we expect more than 1000 participants.

During these last 10 years we have watched and had an impact on the evolution of our field. After 10 years it is essential that the community questions how far we have come and also that it envisions roadmaps and streamlines the trends for the next 10 years, even if we will have to revise them at the next LREC, in two years time.

It is clear that during the last 10 years, our major challenge, as ELRA and the partners from funding agencies, was to ensure that produced resources with (even partial) public funding were properly archived and made available to the different right holders. This concern was not systematically

addressed before the setting up of data centers like LDC and ELRA: a very passive archiving role was taking place! The community managed to get out of this context and succeeded in cataloguing databases and making them widely available through data centers and very often, under fair licensing and pricing schemes.

The emergence of the Web has largely contributed to promote the sharing of language resources but also, like everything else around the web, it has made it critical to excogitate about the nature and the quality of what one gets on the web and the various implications related to such availability (quality, rights and licensing, re-use and exploitation,...). Let me mention just one issue that will be largely debated in this conference: the legal aspect. A survey of the existing licensing schemes conducted by ELDA recently showed that there are more than ten license models, several of them backed up by large institutions, and which handle similar, or even identical, legal matters. Several of them are derived from the software industry traditions and simply transposed to Language Resources. Such abundance and proliferation, instead of helping users to better understand their rights, duties, responsibilities, and liabilities, are confusing most of them who are not prepared to invest time and money in legal advises (though ELRA offers such assistance for free to its members).

Other aspects will also be discussed that may have a substantial impact on the next decade of Language Resources and Evaluation. The proliferation of resources will also be of major importance in our field. It is becoming natural to search the web for the latest dictionary, corpus, list of words, broadcast recordings, etc, free of charge and copyright free. One may identify a large number of such items for a given language. In many cases, there are serious doubts about the quality but also about the origin as well as the actual right owners. It is a challenge for Data Centers like ELRA, LDC, GSK, Chinese-LDC, and similar agencies, to ensure that such trend is fully supported and streamlined for the benefits of all.

The evaluation of technologies in Europe has become a common process in many areas, conducted by many institutions. Some of them specializing in evaluations and devoting all their efforts to such work, others are key players in a given area and volunteer to take over such task to help the community assess the performance of existing technologies. In the later scenario, most of the institutions play a neutral and undisputed role making sure that their own systems or applications are not part of the competition. In both scenarios, the evaluation work is hardly funded and appeals to volunteers who have to carry such work in addition to their regular research. ELRA has always argued for an institution, or a network of centers, at the European level, to play such role, learning from successful experiences of the USA, Japan, China, etc. For the last decade, ELDA managed an evaluation infrastructure and conducted evaluations in more than 20 technologies and is discussing with other partners how to join forces to achieve such goal. This is one of our major challenges for the next couple of years.

ELRA and ELDA have also learned a lot from the science and technology that were presented at the various LRECs. Some of these have to do with ELDA's daily activity. It is good that the infrastructure, set up in 1995, gets seriously challenged at various levels today.

The core activity of ELDA is identification of Language Resources, negotiation of distribution rights and cataloguing such LRs on our online catalogue. Over the years we have revised our catalogue to account for new types of resources (e.g. multimodal), addition of a new section on evaluation packages, etc. Our distribution work is permanently reviewed in the light of the new trends and new distribution mechanisms; in particular today's licensing schemes encourage us to investigate new modes inspired from "open source" principles. ELRA started an initiative to identify all existing Language Resources and make sure that this information is available to its members. We refer to such initiative as the Universal Catalogue. The idea is to offer interested members information about potential resources that may end up in the ELRA Catalogue (if ELRA obtains distribution rights, although this is not guaranteed) or, at least, make our members aware of such existence if they need to discuss bilateral agreements and partnerships with the right holders. Other similar initiatives exist at other institutions and this is yet another reason to offer a forum for discussion like LREC where needs, trends, expectations, plans, partnerships, co-operations can be discussed and brought one step forward.

For ELDA, its strong involvement in the organization of LREC is part of the core mission of the task force set up by ELRA to conduct its strategy and actions. Since the beginning, ELDA considered that it was of paramount importance to join forces with other partners to organize LREC. I would like to take this opportunity to thank them all for their continuous and friendly support. The list is updated on our LREC web site.

Since the beginning of LREC, back in 1998, ELRA has also made sure to meet with representatives of its members. This has been achieved, for instance, through its special offer to its members, who benefit from favorable registration conditions. With more than 150 participants from organizations that are members of ELRA, such goal has been achieved.

As usual let me say a couple of words about ELRA and ELDA. The European Language Resources Association (ELRA, a not for profit European association) was founded in 1995, with the support of the European Commission. The main mission of the Association was to provide a clearing house for language resources, while promoting HLT more generally. In parallel, ELDA, the Evaluations and Language resources Distribution Agency, ELRA's operational body and distribution agency, was created under the instigation of ELRA's board, to handle every activity in relation to the identification, collection, production, marketing and distribution of language resources, along with the participation in HLT evaluation campaigns and other HLT projects, at the European and international levels.

The collection and distribution of language resources are major activities for both ELRA and ELDA and highlight the central role played by both bodies for the advances in the field, however, other crucial services related to language resources and language technologies are also offered. These include the validation of language resources, the production of language resources, and the evaluation of speech and language technologies, with strong involvement in evaluation campaigns to ensure that evaluation resources (data test suites, protocols, methodologies, results, etc.) are packaged and made available to the HLT community, on the model of LR distribution.

If you would like to learn more about ELRA and ELDA, the ELRA/ELDA staff is at your disposal during the conference. You will also find more information on our web sites, at www.elra.info and www.elda.org.

ELRA and the Organizing Committee have made every effort to ensure the success of LREC 2008, making it a fruitful and enjoyable event. We hope that you will enjoy LREC 2008, benefiting at most from the conference programme, the satellite workshops, and the social programme. Hopefully, you will get the opportunity to have a look around, and enjoy Marrakech and Morocco.

Looking back to Granada, I sense a profound and moving story which puts us always in the heart of impressive fames of History and civilizations. Like with the previous locations, in Marrakech, LREC themes will be again in very pleasant company. Marrakech was designed and constructed by the Almoravid Amazigh Caliph Yusuf ibn Tashfin, a king who ruled following the advise of an exceptional wife, Nazha Nafzawiya, who was not only a visionary but an erudite and an instructed lady, which was not very common at that time. The Koutoubia, completed under the reign of another dynasty and Caliph (Yaqub al-Mansur) was named with reference to the surrounding librarians and sellers of manuscripts. I am convinced that you will find by yourself more similarities between Marrakech, Granada, Athens, Las Palmas, Lisbon, and Genoa. It is an easy setting to contemplate thousands of years of History while we are thrown right into the heart of our own stories.

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

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 Marrakech, welcome to LREC 2008.

Khalid Choukri
ELRA CEO and ELDA Managing Director