

## Introduction of the Conference Chair

Nicoletta Calzolari

I wish first to express to Ms Neelie Kroes, Vice-President of the European Commission, Digital agenda Commissioner, the gratitude of the Program Committee and of all LREC participants for her Distinguished Patronage of LREC 2012.

Even if every time I feel we have reached the top, this 8<sup>th</sup> LREC is continuing the tradition of breaking previous records: this edition we received 1013 submissions and have accepted 697 papers, after reviewing by the impressive number of 715 colleagues. We have accepted 30 Workshops and 10 Tutorials. We can't deny that the field of Language Resources and Evaluation is flourishing more than ever! So far more than 1100 people have already registered. From all these signals we see that LREC continues to be – as many say – “the conference where you have to be and where you meet everyone”.

The high acceptance rate is an important characteristic that is at the essence of LREC. We consider of utmost importance to provide a global view of the current trends, in all their dimensions and as reflected in many languages. Multilingualism is a core feature of the field and it is important to show how the field is evolving not only with respect to new methodologies and algorithms but also with respect to more advanced treatment of more and more languages. This is also a strategic choice underlying the importance of safeguarding the heritage of world's linguistic diversity.

In the preparation of the program, while trying to arrange all the pieces of the big puzzle, it is a pleasure touching the hot topics of these days. It is always extremely interesting to spot the major changes with respect to previous editions and monitor the evolution of the field.

Major trends, i.e. not the most crowded topics but those increasing with respect to last LREC:

- “Data”, as normal in our conference, but I would say even more than before: data/corpora (in all the modalities and for many purposes: annotation, extraction, classification, translation, and so on) receive even more attention than last time.
- Machine Translation and multilingualism is a major topic, on which a lot of work is carried out.
- Infrastructural initiatives, strategies, national and international projects are a big issue, as usual inside the LREC community.
- Tools and systems for text analysis at many levels are presented in many papers.
- Temporal and spatial information is also relatively increasing.

New entries, i.e. topics emerging this year:

- New media (twitter, chats and the like)
- Crowdsourcing
- Child language corpora

More or less stable, at the same level as last year are topics such as:

- Semantics and Knowledge, in all their variations: from annotation of anaphoric information, to ontologies and lexicons, disambiguation, named entities recognition, information extraction, to mention just a few.
- Subjectivity, declined in various nuances: emotions, opinions, and sentiments.
- Dialogue and discourse, with contributions from both the Speech and Text communities.
- Speech and Multimodal resources, tools, systems.
- And finally, evaluation and validation methodologies, as an important part of quite many papers.

A slightly declining tendency seems to be associated with:

- Lexical resources
- Grammar, syntax, parsing

Are these solved issues?

As usual, a distinctive feature of LREC is the emphasis given to infrastructural and strategic initiatives. I consider this a very important characteristic of the Language Resource field, one that ELRA has always supported, and one that deserves some reflection. This is probably due to the recognised fact of the necessity to work on massive amounts of data, moreover multiplied by the many languages, and to the infrastructural nature of language resources. The fact that ours is a data-intensive discipline requires building on each other results if we want to achieve something serious, and requires joining efforts. This recognition has led in the last years to the creation of important infrastructures, such as in Europe META-SHARE and CLARIN-ERIC.

We must now seriously think at how to enter with force into the area of “big data” – and also “open data” – the next frontier for competition and innovation, where we have to cope with very strong and organised communities. But we must also learn from them, in particular how to enable working together in huge experiments with thousands of colleagues cooperating on common objectives. This is, I believe, our next step if we want to address the challenges of big data and achieve the status of a mature science. And this requires networking, collaboration and sharing, of ideas and data. I hope LREC helps going in this direction.

## **LREC Innovations**

We continue also the tradition of introducing some innovation at LREC.

- The *LRE Map*, which started only two years ago, is already established, consulted every day and is used in other major conferences. At this LREC we collected descriptions for more than 1200 resources in more than 200 languages! More details on the Map will be found at the ELRA booth where it is presented.
- The novelty of this year is the *Language Library*. It is an experiment to see if we can set up a platform for collaborative work on the processing (annotation, translation, ...) of language resources. The Library will be presented in the ELRA booth as well. Both the LRE Map and the Library have been conceived as services for the community and are being built by the community itself.
- Also this time we have introduced in the program two “*Special sessions*” on emerging new topics – New Media and Crowdsourcing – with a slot dedicated to general discussion.
- We repeat this year the experience of the *EC Village*, so successful last time. We have even more booths, representing a very large number of EC projects.
- A new insertion of this LREC is the *EC Track* on the second day of the conference, organised by Roberto Cencioni, where some EC projects – and related trends – in the field of Language Resources and Evaluation are presented and discussed. This, together with the EC Village, offers the possibility of getting a comprehensive picture of the EC initiatives in the field.

## **Acknowledgments**

And finally, I wish to express my appreciation to all those who made this LREC possible and hopefully successful.

I first thank the Program Committee members, not only for their dedication in the huge task of selecting the papers, but also for the constant involvement in the various aspects around LREC. A particular thanks goes to Jan Odijk, who has been so helpful in the preparation of the program. And obviously to Khalid Choukri, who is in charge of so many organisational aspects around LREC.

I thank ELRA, which is the promoter of LREC, its own conference.

Furthermore, on behalf of the Program Committee, I thank our impressively large Scientific Committee. They did a wonderful job.

A particular thanks goes to the Local Committee, and especially to Mehmed Özkan (its chair): they have worked hard for many months to find the best solutions to local issues.

I express my gratitude to the Sponsors that have believed in the importance of our conference, and have helped with financial support. I am grateful to the authorities, and all associations, organisations, companies that have supported LREC in various ways, for their important cooperation.

I thank the workshop and tutorial organisers, who complement LREC of so many interesting events. A big thanks goes to all the authors, who provide the “substance” to LREC, and give us such a broad picture of the field.

I finally thank the two institutions that have dedicated such a great effort to this LREC, as to the previous ones, i.e. ELDA in Paris and my group at ILC-CNR in Pisa. Without their commitment LREC would not have been possible. The last, but not least, thanks are thus for: H  l  ne Mazo and Sara Goggi, two pillars of LREC without whose commitment for many months LREC would not happen, and the others who have helped and will help during the conference: Victoria Arranz, C  cile Barbier, Paola Baroni, Roberto Bartolini, Riccardo Del Gratta, Francesca Frontini, Olivier Hamon, Val  rie Mapelli, Vincenzo Parrinelli, Valeria Quochi, Caroline Rannaud, Irene Russo, Priscille Schneller. We have solved together the many big and small problems of such a large conference. They will also assist you during the conference.

Now LREC is yours. You – the participants – are the real protagonist of LREC, you will make this LREC great. So, at the very end, my greatest thanks go to you all. I may not be able to speak with each one of you during the Conference (although I’ll try!). I hope that you learn something, that you perceive and touch the excitement, fervour and liveliness of the field, that you have fruitful conversations (conferences are useful also for this) and most of all that you profit of so many contacts to organise new exciting work and projects in the field of Language Resources and Evaluation, which you will show at the next LREC.

I particularly hope that funding agencies all over the world will be impressed by the quality and quantity of the initiatives in our sector that LREC displays, and by the fact that the field attracts practically all the best groups of R&D from all continents. The success of LREC for us actually means the success of the field of Language Resources and Evaluation.

The tradition of holding LREC in wonderful locations with a Mediterranean flavour continues, and Istanbul is a perfect LREC location! I am sure you will enjoy Istanbul during the LREC week. And I hope that Istanbul will enjoy the invasion of LRECers!

With all the Programme Committee, I welcome you at LREC 2012 in such a wonderful country as Turkey and wish you a fruitful Conference.

Enjoy LREC 2012 in Istanbul!

Nicoletta Calzolari  
Chair

8<sup>th</sup> International Conference on Language Resources and Evaluation