

## Introduction of the Conference Chair and ELRA President Nicoletta Calzolari

I wish to express to Mrs. Irina Bokova, Director-General of UNESCO, the gratitude of the Program Committee, of all LREC participants and my personal for her Distinguished Patronage of LREC 2014. Languages – mentioned in the first article of UNESCO Constitution – have been at the heart of UNESCO mission and programmes throughout its history.

I am also especially grateful to Madame Vigdís Finnbogadóttir, UNESCO’s Goodwill Ambassador for languages and former President of Iceland (1980-1996), first woman in the world elected as head of state in a democratic election, for the continuous personal support she has granted to LREC since our first visit in Reykjavík in 2012. In her name the Vigdís International Centre for Multilingualism and Intercultural Understanding has been established under the auspices of UNESCO to promote multilingualism and raise awareness of the importance of language as a core element of the cultural heritage of humanity. I quote a sentence from a recent interview where she says: “The land—our nature—and language, those are our national treasures”: this tells a lot of why this LREC is in Iceland!

### ***Some figures: all records broken!***

LREC 2014, the 9<sup>th</sup> LREC, with its 1227 submissions, has set a new record! We received 21% more submissions than in 2012. We continue the tradition of breaking our own previous records: out of the 1227 submissions, after the reviewing process by well 970 colleagues, we accepted 745 papers. We also accepted 22 workshops and 9 tutorials. More than 1100 participants have already registered at the beginning of May.

These figures have a meaning. The field of Language Resources and Evaluation is continuously growing. And LREC continues to be – as many say – “the conference where you have to be and where you meet everyone”.

Every time I underline the fact that a relatively high acceptance rate (60.7% this time) is for us a reasoned choice. It is important to get a pulse on the situation, to monitor the evolution of the field in the many varieties of approaches and methodologies, and in particular for many different languages. For us, a lexicon in any language is as important as a lexicon in American English. Multilingualism – and equal treatment of all languages – is a feature at the heart of LREC. Other venues promote a sense of exclusivity (also through the equation low acceptance rate and great merit); we always encourage a sense of inclusiveness. This is a typical feature of LREC that makes it a special conference. Quality is not necessarily undermined by a high acceptance rate, but also by the influence of the papers on the community: the ranking of LREC among other conferences in the same area proves this. According to Google Scholar h-index, LREC ranks 4<sup>th</sup> in Computational Linguistics at a similar level of conferences using much lower acceptance rates, just like the LRE Journal also ranks 4<sup>th</sup> in the general field of Humanities, Literature and Arts.

### ***LREC 2014 Trends***

Language Resources (LRs) being everywhere in Language Technology (LT), LREC is a perfect observation point of the evolution of the field. Looking at all the topics, while building the program and putting all the pieces together, the most striking (even if not surprising) new trend was for me the application of sentiment/opinion discovery/analysis to social media shown by so many papers.

A very rough sketch of LREC 2014 major topics and trends, from my viewpoint, is the following:

- There is a completely new topic:
- Linked Data, also the hot topic of this edition
- Topics that were quite new in 2012 and are now consolidated:
- Social Media, in particular combined with subjectivity, as said above
- Crowdsourcing and Collaborative Construction of LRs
- Other increasing (not the biggest in absolute terms) topics with respect to last LREC are:
- Subjectivity: Sentiments, Emotions, Opinions
- Less-resourced languages, in line with the value we give to safeguarding world's linguistic diversity
- Extraction of Information, Knowledge discovery, Text mining: always a very hot topic
- Computer Aided Language Learning
- Stable Big topics:
- Infrastructural issues and Large projects, and also Standards and Metadata, receive the usual attention by the LREC authors
- Lexicons and Corpora (i.e. the most typical “data”), of many types, modalities and for many purposes and applications: they are the prominent and most crowded topic
- Semantics and Knowledge, in all their variations: from annotation of anaphoric information, to ontologies and WordNets, sense disambiguation, named entities recognition, information extraction, to mention just a few
- Syntax, Grammar and Parsing continues to be a largely represented topic: not solved
- Machine Translation and Multilingualism are areas on which a lot of work is carried out
- Speech and Multimodality keep the same level: good but not enough
- Dialogue and discourse, with contributions from both the Speech and Text communities
- Evaluation is pervasive/everywhere: we are proud to give evidence to its being an essential feature in the LT landscape
- Tools, systems for text analysis and applications are presented in many papers

A usual observation is the relevance of *infrastructural issues* and the attention that LREC – and ELRA – pay to them. They are mostly neglected in other conferences. Infrastructural issues play an important role for the field of LRs and for the LT field at large. But it is a fact that the first to recognise their importance have been people of the LR area. LRs are themselves of infrastructural nature and quite naturally call for attention to these issues. The infrastructural nature of LRs, captured by the term “Resources”, was highlighted in the Introduction of Antonio Zampolli to the 1<sup>st</sup> LREC in Granada in 1998.

The fact that so many topics are represented at LREC means also that all the various LR and LT sub-communities are present at LREC: this increases the LREC impact and gives to LREC the characteristic of being a true melting pot of cultures, and an enabler of new cooperation initiatives.

### ***15<sup>th</sup> LREC Anniversary***

LREC was born in 1998 and on the occasion of its 15<sup>th</sup> Anniversary, Joseph Mariani has prepared an analysis of all the past LREC Proceedings, rediscovering the dynamism of the field while looking at the major contributors, topics, trends, also comparing them with an analogous survey done for the speech community on the Interspeech conference series. There will also be a Quiz for all the LREC participants and a winner! The survey paper is in the Proceedings as a special paper for the 15<sup>th</sup> Anniversary and will be presented at the Closing Session.

### **ELRA and LREC: a tradition of innovations at the service of our community**

I am proud to announce a number of recent initiatives of ELRA and LREC that touch topics that are at the forefront of a paradigm shift and together help advance our field and increase confidence in scientific results. As an introduction I use some words of Zampolli in 1998: “The need to preserve, actively promote the use of, and effectively distribute LR, has caused the USA and EU authorities to put in place, respectively, LDC (the Linguistic Data Consortium) and ELRA (the European Language Resources Association)”, observing also that their activities “demand regular updating to reflect technical and strategical evolution of their environment”. We try to keep with this recommendation.

These innovations – introduced by ELRA and /or LREC – must not be seen as unrelated steps, but as part of a coherent vision, promoting a new culture in our community. We want to encourage also in the field of LT and LRs what is in use in more mature sciences and ensure reproducibility as a normal part of scientific practice. We try thus to influence how our science is organised or should be organised in the future.

I give here a quick picture of some innovations that are critical for the research process and constitute a sort of manifesto for a new kind of sustainability plan around LRs.

### ***LRE Map***

The LRE Map (<http://www.resourcebook.eu/>), started in 2010, is now an established tool, consulted every day and used in other major conferences. At this LREC we have collected by the authors descriptions for more than 1000 resources in more than 150 languages!

Spreading the LR documentation effort across many people, instead of leaving it only in the hands of the LR distribution centres, we also encourage awareness of the importance of metadata and proper documentation. Documenting a LR is the first step towards identifiability, which in its turn is the first step towards reproducibility.

Recognising the value of Linked Data, we just published the LRE Map in LOD (Linked Open Data).

### ***Share your Language Resources and Reproducibility of research results: the vision***

After encouraging sharing LR metadata, the next step is sharing the actual content. ELRA has embraced in the last years the notion of “open LRs”: we show this also with the “Share your Language Resources” initiative started in this LREC. With it we ask all the authors to consider making background data available with their paper. More than 300 LRs have been made available: a big success for the first experiment! Showing the community commitment to sharing.

LRE Map and Share your LRs must be seen not as isolated initiatives, but as complementary steps towards implementing a new vision of the field. On one side we encourage opening data that could be valuable to others, on the other we try to encourage a sort of cultural change in our community.

Here the *vision*: It must become common practice also in our field that in conferences and journals when you submit a paper you are offered the opportunity to upload the LRs related to your paper. We must unlock the material that lies behind the papers: the adoption of such a policy will make the whole picture clearer. We had to fight in the ‘90s for concepts like “reusability”, which finally led to promoting the need of developing standards in our field (this was still a hot topic in 1998 at the time of the 1<sup>st</sup> LREC). Now the need for standards is consolidated and we consider it normal, but we need to start another campaign for encouraging more resource sharing. Researchers are not yet sharing very well; they tend to hold back knowledge. I hope that this sharing trend will be more easily embraced by younger colleagues who are familiar with everyday use of social media of all sorts and free ideas sharing: we must port the same attitude in the research environment. This will fundamentally change the way of making science, in a sort of light revolution towards openness of science in all its facets. Hopefully it will diminish the unfortunate phenomenon of reinventing the wheel from time to time, instead of building on your colleagues’ findings.

This vision has to do with many important aspects: shifting to a culture of sharing, re-use, reproducibility of research results. If we want to become a mature science we should make data sharing become “normal” practice. Even more important in a data-intensive discipline like LT. The small cost that each of us will pay to document, share, etc. should be paid back benefiting of others’ efforts and become worthwhile. This will also lead to a greater opportunity of collaboration, encouraging bigger experiments by larger collaborative teams (something else we should learn from more mature sciences). Moreover, reproducibility encourages trust.

### ***ISLRN***

A major achievement of ELRA has been the recent establishment of the *International Standard Language Resource Number* (ISLRN) (<http://www.elra.info/Establishing-the-ISLRN.html>). It is a unique identifier to be assigned to each LR. Organised and sustained by ELRA, LDC and AFNLP/Oriental-COCOSDA, the ISLRN Portal provides unique identifiers to LRs. LRs in the ELRA

and LDC catalogues have been the first to get an ISLRN (just one if a LR is stored in both catalogues!).

When you publish a LR it can get an ISLRN and thus become a citable product of research. Data/LR citation must become normal scientific practice also in our field, as it is in others. To make a LR citable can then pave the way to the design of a sort of “impact factor” of LRs. This can become an important incentive for the field, so that researchers can get the credit they deserve also for the LRs they developed.

ISLRN is not only linked to the possibility of getting proper “recognition” for LR developers. It would also enhance experiment replicability, an essential feature of scientific work. It may thus become a very important advance in our field.

#### ***META-SHARE sustainability by ELRA***

Through these initiatives we try to encourage community efforts towards: documentation of LRs, possibility of identification of LRs, LR sharing, making research results reproducible. There is a lot of buzz these days around these types of topics. As I said above, all these initiatives are closely related and must become integrated with each other.

For them to become common research practices these activities must be well organised and require good mechanisms behind to become possibly a set of related services on a common platform. Pooling together data from all the research described in conference and journal papers will obviously need an infrastructure for distributing research results and such a LR platform must be sustained.

The ELRA Board has decided to support the META-SHARE platform, but META-SHARE – as sustained by ELRA – must in turn be adapted to be able to support these types of initiatives and thus become also a platform for sharing reproducible research results. We must find ways to make these practices as easy as possible and rewarding for the researcher. META-SHARE – in ELRA view – should become also the obvious repository (recognised by the community) where all these types of actions are sustained and where all research results become available, discoverable, identifiable, and citable. ELRA is taking these steps to start enabling to keep track of connected research activities like papers and supporting underlying resources, in an all-inclusive way.

#### ***LREC Proceedings in Thomson Citation Index***

A great recent achievement for ELRA and LREC has been the fact that the LREC 2010 and LREC 2012 Proceedings have been accepted for inclusion in CPCI (Thomson Reuters Conference Proceedings Citation Index). This is a significant achievement for LREC and it will provide all LREC authors with a deserved recognition. It is for us of great satisfaction, in particular for the benefit it can bring to young colleagues.

#### ***ELRA 18<sup>th</sup> anniversary and NLP12***

Coordination is an important issue when infrastructural issues are at stake. None of the actions above can or should be conducted and tackled in isolation.

For this reason we – ELRA – organised, on the occasion of ELRA majority as its 18<sup>th</sup> anniversary, the first meeting of the major associations/organisations in the field of Language Resources and Technologies, Computational Linguistics, Spoken Language Processing, Big Data and Digital Humanities, the so-called NLP12 (<http://www.elra.info/NLP12-Paris-Declaration.html>). We started to discuss issues of common interest to coordinate some of the activities and we adopted some common resolutions, such as the encouragement of language resources and tools sharing and promotion of best practices for language resource citation in publications.

Together we should be able to take the necessary steps to better serve the field and the respective communities and to strengthen the bridges between various communities (e.g. Language Technology and Humanities).

### ***ELRA for Open science***

I am excited and proud that we – as ELRA and LREC – can contribute to such a (quiet) revolution towards shaping a new type of *open scientific information space* for the future of our field, the Language Resources and Technology future. I have always felt it is our duty to use the means that we have in our hands to try to shape the future of the field, and in this case to play a role in how to change scientific practice and have an impact on the overall scientific enterprise!

Trying to be always forward-looking and to act in a proactive way to serve the field, ELRA continues to be a community-aware association. I would like to work for it to become more also a community-driven association. We would like to discuss with all those who are interested about how to tackle the challenge of truly open research (which is more than open access!) so that we can take the necessary further steps to make this process more efficient, faster and more collaborative.

It is clear that in such a campaign for the cause of reproducibility and open science and for a proper system of attribution and citation – two closely related aspects– we must involve also funding agencies that should help in supporting the necessary policy actions. For sure we will involve in this initiative the NLP12 group. But I strongly believe that the most important change must come from the mind-set of researchers. This is where LREC can help, I hope ...

The message that ELRA has for its community, the LREC community, is: We are here to help!

### **Acknowledgments**

In this last part I wish to express my deepest gratitude to all those who made this LREC 2014 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. To Joseph Mariani for his always wise suggestions. And obviously to Khalid Choukri, who is in charge of so many aspects around LREC.

I thank ELRA and the ELRA Board: LREC is a major service from ELRA to all the community! A very special thanks goes to Sara Goggi and H el ene Mazo, the two Chairs of the Organising Committee, for all the work they do with so much dedication and competence, and also the capacity to tackle the many big and small problems of such a large conference (not an easy task). They are the two pillars of LREC, without whose commitment for many months LREC would not happen. So much of LREC organisation is on their shoulders, and it is visible to all participants.

A particular expression of gratitude goes to the Local Committee, and especially to Eir kur R ognvaldsson (its Chair) and Sigr un Helgadóttir: they have worked with great commitment and enthusiasm for many months for the success of LREC always looking at the best solutions to the many local issues.

All my appreciation goes also to the distinguished members of the Local Advisory Board for their constant support.

Among the Icelanders I wish to mention Gu r un Magn sadóttir, for a very simple reason: the idea of having LREC in Iceland came out during a lunch that the two of us had together in Berlin!

I express my gratitude to the Sponsors that believe 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. Furthermore, on behalf of the Program Committee, I praise our impressively large Scientific Committee. They did a wonderful job.

I thank the workshop and tutorial organisers, who complement LREC of so many interesting events.

A big thanks goes to all the LREC 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 ILC-CNR in Pisa. Without their commitment LREC would not have been possible. The last, but not least, thanks are thus, in addition to H  l  ne Mazo and Sara Goggi, to all the others who have helped and will help during the conference: Victoria Arranz, Paola Baroni, Roberto Bartolini, Irene De Felice, Riccardo Del Gratta, Francesca Frontini, Ioanna Giannopoulou, Johann Gorlier, Olivier Hamon, J  r  my Leixa, Valerie Mapelli, Vincenzo Parrinelli, Valeria Quochi, Caroline Rannaud, Irene Russo, Priscille Schneller. You will meet most of them during the conference.

I also hope that funding agencies will be impressed by the quality and quantity of initiatives in our sector that LREC displays, and by the fact that the field attracts 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.

And lastly, my final words of appreciation are for all the LREC 2014 participants. Now LREC is in your hands. You are the true protagonist of LREC; we have worked for you all and you will make this LREC great. I hope that you discover new paths, that you perceive the ferment 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.

LREC is not exactly in a Mediterranean location this time, even if all the hot water around gives some Mediterranean flavour! But the tradition of holding LREC in wonderful locations continues, and Reykjav  k is a perfect LREC location! I am sure you will like Reykjav  k and the friendliness of Icelanders. And I hope that Reykjav  k will appreciate the invasion of LRECers!

With all the Programme Committee, I welcome you at LREC 2014 in such a wonderful country as Iceland and wish you a fruitful Conference.

Enjoy LREC 2014 in Reykjav  k!

Nicoletta Calzolari

Chair of the 9<sup>th</sup> International Conference on Language Resources and Evaluation and ELRA President