

Introduction of the Conference Chair and ELRA President Nicoletta Calzolari

Welcome to the 10th edition of LREC in Portorož, back on the Adriatic Sea!

I wish to express to his Excellency Mr. Borut Pahor, the President of the Republic of Slovenia, the gratitude of the Program Committee, of all LREC participants and my personal for his Distinguished Patronage of LREC 2016.

Some figures: previous records broken again!

It is only the 10th LREC (18 years after the first), but it has already become one of the most successful and popular conferences of the field. We continue the tradition of breaking previous records. We received 1250 submissions, 23 more than in 2014. We received 43 workshop and 6 tutorial proposals.

The Program Committee is confronted at every LREC with a harder and harder job, going through 3750 reviews, to understand – also beyond the scores and in particular when they greatly differ – the relevance, the novelty, but also the appropriateness for an oral or poster presentation. We have in the program 744 papers: 203 Orals and 541 Posters.

We recruited an impressive number of reviewers, 1046 (76 more than in 2014), to keep the number of papers per reviewer rather low. This was a great effort in which a very large amount of our community was involved. To reach this number we had to invite 1427 colleagues, out of which 182 declined and 199 regrettably did not answer. At the end I must say that a number of reviewers (not so many) did not do their duty and we had to recruit some others in a hurry.

We also have 30 Workshops and 6 Tutorials.

More than 1100 participants have already registered at the beginning of May.

These figures and the continuously growing trend have a clear meaning. The field of Language Resources and Evaluation is very alive and constantly flourishing. And LREC seems still to be – as many say – “the conference where you have to be and where you meet everyone”.

LREC acceptance rate: a reasoned choice

Also this time, as I usually do, I want to highlight the LREC acceptance rate, 59.52 this year, unusual in other major conferences but for us a reasoned choice. This level of acceptance rate is a special feature of LREC and is probably one of the reasons why LREC succeeds to give us an overall picture of the field and to reveal how it is evolving. For us it is in fact important not only to look at the top methods but also to see how much various methods or resources are able to spread, for which purposes and usages and among which languages. Multilingualism – and equal treatment of all languages – is an essential feature of LREC, as it is the attempt of putting the text, speech and multimodal communities together.

The acceptance rate goes together with the sense of inclusiveness that is important for us (instead of the sense of elitism associated with a low acceptance rate).

And I want to underline again that quality is not necessarily undermined by a high acceptance rate, but it is also determined 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 4th in Computational Linguistics top publications.

I was really proud when a colleague recently told me that LREC, with its broad variety of topics, is the conference where he gets more ideas than in any other!

LREC 2016 Trends

From one LREC to another I have tried (since 2004) to spot, even if in a cursory and subjective way, the major trends and the going up and down of certain topics.

After highlighting the major trends of 2016, also in comparison to 2014 and previous years, I make here few general considerations. The comparison with previous years highlights the topics where we find steady progress, or even great leaps forward, the stable ones and those that may be more affected by the fashion of the moment.

Trends in LREC2016 topics, also compared to 2014

Among the areas that continue to be trendy and are increasing I can mention:

- Social Media analysis, started in 2012 and increasing in 2014, is doubling again
- Discourse, Dialogue and Interactivity
- Treebanks, with a big increase with respect to the past
- Less-resourced languages
- Semantics in general and in particular Sentiment, Emotion and in general Subjectivity
- Information extraction, Knowledge discovery, Text mining
- Multilinguality in general and Machine Translation
- Evaluation methodologies

Unsurprisingly many papers in the “usual” topics:

- Lexicons, even if a bit decreasing
- Corpora
- Infrastructural issues, policies, strategies and Large projects: topics that receive special attention at LREC, differently from other major conferences. Another distinguishing feature for LREC.

Newer trends:

- Digital Humanities
- Robotics

Stable topics:

- Speech related topics, a little increasing but not as much as we would like
- Multimodality
- Grammar and syntax
- Linked data, a new topic in 2014, remains stable
- Computer Aided Language Learning, an increasing topic in 2014, is stable

Less-represented topics with respect to the past:

- Web services and workflows
- Sign language (probably because there is a very successful workshop on this)
- Ontologies
- Standards and metadata
- Temporal and spatial annotation
- Crowdsourcing

Overall trends, from 2004 ... and before

From 2004 we observe a big increase in papers related to Multilingualism and Machine Translation. This may also be related to the funding of Machine Translation projects from the European Commission.

The analysis of Social media and Subjectivity with sentiment, opinion, emotions, has started in this time span and is not only well consolidated but also continually expanding.

There is declining tendency for papers related to grammar and syntax. This however makes even more interesting the high increase of papers on Treebanks this year.

There seems to be a small decrease of papers on Lexicons and Lexical acquisition as well as on Terminology. It was probably a more popular topic years ago when many WordNet and FrameNet lexicons were built for many languages.

I have recently been reminded by a colleague at ILC of a paper I wrote many years ago with some considerations on Computational Linguistics as reflected by the papers at COLING 1982. There was obviously no mention of Language Resources (the term itself was coined by Zampolli later on), but I underlined already then the element of novelty constituted by the area that was called at that time “Linguistic Data Bases”, with some papers on dictionaries “in machine-readable form”. While I never mentioned in my review, and probably the papers too, the word “Corpus”! Only 30 years ago Computational Linguistics was a totally different field. The new area of Language Resources was born some years after those initial pioneering sparse papers. But this new topic, as testified by the success of LREC, has expanded incredibly fast.

And a new community has taken shape around Language Resources. A peculiarity of this community is the attention paid to infrastructural issues, to overall strategies and policies. This is also due, I believe, to the fact that in many cases we have to work in large groups, for many languages, we must be able to build on each other work, to connect different resources and tools, to make available what already exists and use standardised formats. Infrastructures (on many dimensions) are really needed for this field to progress.

I wrote in the introduction to LRE2006, 10 years ago: “Do we have revolutions? Probably not. Even if the stable growth of the field brings in itself some sort of revolution. After a proliferation of LRs and tools, we need now to converge. We need more processing power, more integration of modalities, more standards and interoperability, more sharing (in addition to distribution), more cooperative work (and tools enabling this), which means also more infrastructures and more coordination.” I think that many of the needs that I expressed then are being achieved today, as testified by the papers in this edition of LREC, and therefore we can probably speak of a sort of quiet revolution.

LREC Proceedings in Thomson Citation Index

I remind also that since 2010 the LREC Proceedings have been accepted for inclusion in CPCI (Thomson Reuters Conference Proceedings Citation Index). This is quite an important achievement, providing a better recognition to all LREC authors and useful in particular for young colleagues.

ELRA and LREC

ELRA 20th Anniversary: achievements, promotion of new trends and community building

In 2015 we organised a workshop for the 20th anniversary of ELRA, founded in 1995. I think it is a big success the fact that ELRA has remained in the Language Technology picture with growing influence in these 20 years, even more so given that ELRA does not rely on specific public funding.

ELRA has implemented, over the years, many services for the language resource community, promoting the integration not only of different modalities but also of different communities. And ELRA itself has evolved to reflect the evolution of the field and sometimes to anticipate new trends.

I mention here a few services: Evaluation, Validation, Production of LRs, Networking, LREC, LRE Map, Sharing of LRs, Licensing wizard, ISLRN, LRE Journal, Less-resourced languages committee. These initiatives must not be seen as unrelated steps, but as part of a coherent vision, promoting a new culture in our community. Many of ELRA actions are of infrastructural nature, knowing that research is affected also by such infrastructural activities.

LREC is probably the most influential ELRA achievement, with the major impact on the overall community. Also through LREC, ELRA has certainly contributed to shape our field, making the Language Resource field a scientific field in its own right.

This is testified also by the LRE journal, coedited by Nancy Ide and myself, with submissions in continuous increase.

Citation of Language Resources

This year we introduced and encouraged citations of Language Resources, providing recommendations on how to cite: we have in fact added a special References section. This must become normal practice, to keep track of the relevance of LRs but also to provide due recognition to those who work on language resources.

This will be implemented also in the LRE journal.

Replicability of research results

We continue to promote, also through many of the initiatives above, greater visibility of LRs, sharing LRs in an easier way and replicability of research results as a normal part of scientific practice. ELRA is thus strengthening the LR scientific ecosystem and fostering sustainability.

Acknowledgments

As usual, it is my pleasure to express here my deepest gratitude to all those who made this LREC 2016 possible and hopefully successful.

I first thank the Program Committee members, not only for their dedication in the huge task of selecting the papers, the workshops and tutorials, 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 programme. 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 Organizing 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 this is visible to all participants.

I am grateful also to the Local Committee, especially to Simon Krek (its Chair) and Marko Grobelnik, for their help in organising a successful LREC.

My appreciation goes also to the distinguished members of the Local Advisory Committee for their constant support.

I express my great 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 the associations and organisations that have supported LREC in various ways.

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 always dedicate a great effort to LREC, i.e. ELDA in Paris and ILC-CNR in Pisa. Without their commitment LREC would not be possible. The last, but not least, thanks are thus, in addition to H el ene Mazo and Sara Goggi, to all the others who – with different roles – have helped and will help during the conference: Paola Baroni, Roberto Bartolini, Dominique Brunato, Irene De Felice, Riccardo Del Gratta, Meritxell Fern andez Barrera, Francesca Frontini, Lin Liu, Val erie Mapelli, Monica Monachini, Vincenzo Parrinelli, Vladimir Popescu, Valeria Quochi, Caroline Rannaud, Irene Russo, Priscille Schneller, Alexandre Sicard. 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 2016 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 again in a Mediterranean location this time! I am sure you will like Portorož and Slovenia and the Mediterranean atmosphere. And I hope that Portorož and Piran will appreciate the invasion of LRECers!

With all the Programme Committee, I welcome you at LREC 2016 and wish you a very fruitful Conference.

Enjoy LREC 2016 in Portorož!

Nicoletta Calzolari

Chair of the 10th International Conference on Language Resources and Evaluation and ELRA President