

LREC2006: Introduction of the Conference Chair

This is the fifth edition of LREC, which means that LREC is only 8 years old, not even a decade, but many things have changed in these few years.

In 1998 Antonio Zampolli understood that a new community was forming, around the topic of Language Resources (LRs) and Evaluation, a community whose interests were not served completely by the major conferences of the area of Computational Linguistics. His intuition, like many others before, proved to be absolutely right, as confirmed by the ever growing number of submissions to LREC and by its extremely large attendance. When LREC was established in 1998, LRs – and with them Evaluation – were starting to receive by larger sections of the HLT (Human Language Technology) community the attention that for many years was given to other aspects of language technology. LREC has already become, after just 8 years, a ‘traditional’ and very big conference in the sector of Computational Linguistics.

What does it mean? It is a confirmation that LRs constitute indeed the necessary infrastructure for any Language Technology (LT) and Evaluation project. This was the great intuition of Antonio and of some of us (the oldest here) back in the late ‘80s. Among these I’d like to mention also Don Walker, who played an important role in making the role of LRs recognised within the Computational Linguistics community.

The ‘data-driven’ approach is no longer something for which to fight, as it was many years ago for colleagues like Geoffrey Leech: some of us still remember how his corpus analyses were badly received at an European ACL of the ‘80s. This era seems so far today, and the youngest may consider it absurd.

Statistical methodologies are now by far the major trend in computational linguistics, even too much, sometimes at the expenses of serious linguistic analyses. In the same direction, robustness is of major relevance for the production of effective applicative systems. And data, i.e. LRs, are behind these trends. We have to pay attention to avoiding that innovative and valuable trends do not become just ‘fashions’.

At the same time the recognition of the need for good quality, for comparing results, for measuring progress, and so on, has given more and more importance to evaluation methodologies, as we all know.

LREC remains the best observatory for an examination of the evolution of the field of LRs and Evaluation, and by consequence of LT. Looking retrospectively at the various LRECs, and at this LREC now, we can – and maybe must – ask ourselves a few questions:

- i) whether, how and how much LRs have influenced the evolution of LT,
- ii) how the field of LRs itself is changing,

and based on these, but more critically for our future, questions such as:

- iii) how the achievements of the last years must influence our future directions of research,
- iv) if completely new trends are in front of us,
- v) what will be the role of LRs in the future of LT,
- vi) which infrastructural, strategic, cooperation or coordination initiatives are needed in the next years for a better development of the field.

Just a few words on the first two complementary points.

It is the merit of LRs (or at least a big part of the merit) if LT is changing so much, is acquiring maturity, and is gradually attaining the robustness needed to become truly useful in real world

applications. This is probably the biggest effect of LRs, causing also a big transformation of LT, from 'just' a R&D sector to a technology with a great impact in the society.

But also the field of LRs is changing in many ways, and consequently the needs of our community are different. It is more mature, which is a trivial observation, but this may have not trivial consequences. At the beginning of the '90s three major areas were perceived – and described probably for the first time by Antonio and me at a workshop in Santorini in 1993 – as critical for the development of the field:

- i) standardisation of LRs,
- ii) creation of basic LRs and their annotation,
- iii) distribution of LRs.

Major projects and initiatives of the '90s had objectives related to the implementation and satisfaction of these needs: i) standards were defined, accepted and used; ii) many large LRs were created. iii) ELRA came out from this vision.

Where we are with respect to these needs? Work on the three tracks is still going on, as we see from LREC papers. A lot has been achieved, but a lot still has to be done, both in quantity (more standards, more LRs for more languages, the web considered an invaluable source, what was considered large 10 years ago is no longer large, more distribution) and in new ways of approaching the problems. A fourth area has clearly acquired an increasingly larger relevance for LRs and LT, i.e.:

- iv) methods for automatic acquisition of linguistic (or other) information.

But what is of interest to me is:

- are the three/four areas still valid, or the most critical, today ?
- in which ways these areas are changing?
- which are the new needs of the field?

I think LREC is helpful in answering these and similar questions.

Preparing the programme of this Conference what have I noticed? Let me quickly touch just few of the issues, and of the questions raised by the set of submissions.

If we consider the traditional levels of linguistic analysis, morphology seems no longer a central issue. It is probably almost solved, and also for syntax there is a lot of consolidation of achieved results. While semantics is still a topic for research and development, it is really at the centre of the scene, and this happens looking both at works on corpora and on lexicons. Ontologies are becoming central.

Systems maintain their importance also in a conference for LRs and evaluation, in particular for information extraction, information retrieval, machine translation, question answering,

Many papers, more and more, focus on evaluation, either as evaluation of tools, systems or also of LRs themselves (validation in this case). Many also on evaluation methodologies per se and on usability and user satisfaction.

Moreover, new topics are emerging, linked to subjectivity more than to the 'objective' aspects of meaning, and interestingly this happens both for spoken and written research. I mean topics such as discovery, analysis, representation of sentiments, affect, opinions. This is a new area of research with potentially enormous applicative impact, in areas such as business, marketing, intelligence. The interest for these new topics does not exclude that more 'objective' areas do not present challenges, on the contrary. Despite the progress in the ability to semantically annotate texts, we are far from having 'solved' the problem of 'meaning' or of semantic interpretation of texts. To grasp, manipulate, and effectively use content, both objective and subjective aspects of it, remains the big

challenge of our field. Intelligent access to content is thus a goal, maybe a revival – hopefully more successful – of the old Artificial Intelligence with new and more powerful means, i.e. new batteries of tools and resources.

Another hot topic is multilinguality. This has been sometime neglected, while it will be a major unifying factor for future R&D.

The same is true for multimodality, which is more and more important. This emerges not only from the quantity and quality of submissions, but also from two of the satellite workshops.

And general topics such as LR infrastructures and architectures, large projects, organisational and policy issues see a big growth, receiving more and more attention.

Do we have theoretical issues? Or ours is just a practical empirical field? We have both. The ‘data-driven’ approach is by nature empirical, and statistical methods are certainly pervasive, but theoretical reflections on language are imperative also in this area.

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.

Where are we going?

The set of LREC papers, of workshops, tutorials, are together delineating some trends. It’s up to all of us to draw the consequences. In a workshop, the last day of the Conference week, we will try to see together what are the emerging trends, the challenges, the consolidated achievements, the promising new directions, the necessary synergies, the breakthroughs – if any.

This is one of the important roles of LREC, to help the community to reflect on itself to have a better vision of the future. I do not want to draw conclusions here. I leave it to the group of us together at the Roadmap workshop to try to do that.

But I would like to have some reflections on these issues at the next LREC, which may be appropriate after the first decade of its life.

Acknowledgments

First of all my thanks go to the Programme Committee, that at every LREC has a harder work, having to deal with an ever increasing number of submissions, of a better quality, while our structure is still the same. I am sure we have made mistakes, and we ask for your forgiveness.

Also on behalf of all the members of the Programme Committee, I warmly thank all the various Committees that have made this LREC possible, and hopefully successful.

I thank ELRA and the ELRA Board, for their continuous commitment to LREC.

I thank our impressively large Scientific Committee, composed of more than 300 colleagues from all over the world. They did a wonderful job, succeeding to complete their reviews in time for so many papers.

We are also indebted to the International Advisory Board and the Local Advisory Board, that have provided moral support to our Conference.

I am grateful to authorities, associations, organisations, committees, agencies, companies that have supported LREC in various ways, for their important cooperation.

I express my gratitude to the sponsors that have believed in the importance of our Conference, and have helped with economic support.

I specially thank the Local Organising Committee, Lucia Marconi, Paola Cutugno and Daniela Ratti, who had succeeded in finding solutions to local problems, despite having often to face delays or changes in decisions of relevance to local matters. We have solved together the many big and small problems of a large Conference like this. They will assist you during the days of the tutorials/workshops and the Conference.

I thank the workshop, tutorial, and panel organisers, who surround LREC of so many interesting events. A big thanks to all the authors, who provide the content to LREC, and give us such a broad picture of the field.

This time I wish also to thank two institutions which have provided economic support and dedicated so much effort, in term of manpower, to this LREC, as to the previous LRECs, i.e. ELDA in Paris and my institute, ILC-CNR in Pisa and Genoa. Without their dedication LREC would not have been possible.

So I arrive to the last, but not least thanks, dedicated, with all my sympathy, to the people of these institutions who have worked so intensely to make this LREC possible in all its details. Despite the distance (Paris-Pisa) they have worked together as a unique and wonderful team, with enthusiasm and dedication. My biggest thanks go to H el ene Mazo and Mathieu Robin-Vinet at ELDA, while at ILC to the new LREC entries Paola Baroni and Alessandro Enea, to the old LREC staff Vincenzo Parrinelli, Sergio Rossi and in particular Sara Goggi, who have become over the years one of the pillars of LREC. I cannot list the many tasks they carried out, but I can say for sure that without their daily work and real commitment since many months, LREC would not happen.

As I said last time, now LREC is in your hands, the participants. You are the protagonist of LREC, you will make this LREC great (I am sure). So at the very end my biggest thank goes to all of you. I may not be able to speak with each of you during the Conference (I'll try). I hope that you learn something, that you perceive and touch the ebullience, exuberance and liveliness of the field, that you have fruitful conversations (conferences are useful also for this), most of all that you profit of so many contacts to organise new exciting work and programmes in the field of LRs and evaluation, which you will show at the next LREC.

I particularly hope that funding agencies all over the world are impressed by the quality and quantity of 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. This is a sign they must take into account in their programmes and funding strategies. The success of LREC means to us in reality the success of the field of LRs and Evaluation.

With all the Programme Committee, and with the Genoa, Paris and Pisa teams, I welcome you at LREC 2006 in Genoa and wish you a wonderful Conference.

Enjoy LREC!

Nicoletta Calzolari
Chair

5th International Conference on Language Resources and Evaluation