Introduction of the Conference Chair

Nicoletta Calzolari

Antonio Zampolli launched the idea of a Language Resources and Evaluation Conference (LREC) during an ELRA Board meeting. And Angel Martin Municio proposed for the first LREC Granada. It was the perfect combination for a new adventure. Which continued with Athens, Las Palmas, and now Lisbon… Here we are, and Antonio and Angel are with us only in spirit. But they are with us.

Antonio Zampolli understood that, notwithstanding the many conferences, there was space for something different, and a gap to be filled by an event that could gather all the researchers aware of the relevance of Language Resources (LRs) and of Evaluation, as two 'transversal' lanes with respect to all the 'vertical' human language technologies. Not only there was space for it, but, as the success of LREC has proven, there was the need of it.

We dedicate this LREC 2004 to Antonio Zampolli. In particular we have a special plenary session with three of the 'oldest' friends of Antonio speaking to him and for him. I think he will like that.

We owe to Antonio also the term "language resources", as well as the recognition of their infrastructural role for any Human Language Technology application. We must not forget the leading role of a few visionary people for making the data-driven approaches accepted by the scientific community, and Antonio Zampolli was the most important figure to realise that and to push for this vision. Only about a decade ago, around the '80s, it was considered by many colleagues almost a 'shame' to have to deal with data, such a trivial matter! Only methods and algorithms were considered by many scientifically valuable. The problem was that these rule-based methods were often valid for the examples at stake, but not effective for real situations. This was particularly true in the written or textual area, while in the spoken area statistical methods, and therefore data, were recognised as valuable, or even necessary, well in advance.

A revolution has taken pla ce since then. LRs and Evaluation traverse today horizontally every applicative area of HLT, as the '04 edition of LREC very clearly testifies.

If we compare the content of the four editions of LREC, and try to make even a very superficial and cursory analysis of the prominent areas covered in the four conferences, we cannot avoid noticing a number of trends. The field is evolving, and these trends reflect very clearly the evolution of the field and the emerging needs, and provide us with a picture of where our field moves, and how it changes. Just a few quick remarks.

The focus of the attention is moving – on the continuum of the LR space – from one edition to the other: from issues of morphology and tagging, to grammars and treebanks (many in '02), then terminology and knowledge, semantics, semantic web and ontologies, pragmatics, multimodal dialogue, and how to model emotions (there was no paper on emotions in '98).

In this '04 Conference we have such a broad spectrum of *tools, components, systems, applications* represented, that at a first glance it may appear we are at another conference, but they are rightly here. LRs in fact occupy more and more space in our everyday work. Summarisation, question answering, machine translation, speech-to-speech translation, cross lingual information retrieval, information extraction, document classification, automatic indexing of broadcast news, topic detection, text mining, e-learning, to mention just a few, need data, often lots of data. And need

evaluation, and good methods for evaluation. Moreover, it is very recent the recognition of the strategic importance, both in politic and economic terms, of being able to build a new system for a given language in a very short time, or to adapt or tune an existing one very quickly, which crucially depends on the availability of large quantities of data and on the ability to process them. This is another shift of focus, i.e. from papers on 'data almost per se' to 'what we use the data for' and 'how we use the data'. This is an important change of perspective.

Finally, an impressive amount of papers this time are on *'how to acquire data'*, i.e. about methodologies and techniques for machine learning, automatic acquisition and/or classification of information. They aim at creating LRs, and at the same time rely on LRs, at some stage either of implementation or of evaluation. They are the real trend and the challenge of the last years, and one of the most promising research areas for the next years.

The other pillar is *Evaluation,* without which no technology is credible. Many evaluation resources and many evaluation methodologies are presented at this LREC: evaluation in many cases of resources, tools or systems where semantics is at stake, from evaluation of disambiguation systems to ontology platforms, from machine translation to summarisation. Both American and European large evaluation campaigns are well represented. Also validation of LRs themselves acquires more and more importance, as a fundamental step to accompany any distribution activity. Validation is closely linked to standards.

LREC, with its extremely broad spectrum of topics, and even more with its coverage of spoken, written and multimodal LRs, allows an assessment of the *level of maturity* not only of the field of LRs, but of HLT in general, because of the clear interaction between LRs and HLT. This is why it is important to have a conference providing an overview of "what exists", not only of what is new. This has always been an important parameter for evaluation of papers for LREC. LREC is a different type of Conference from ACL or COLING, not better or worse, just different. Also the acceptance rate corresponds to how LREC wants to characterise itself. I stress again here that LREC is a conference where it is important to report not only on what is methodologically new, but also on which LRs exist, for which languages, in which state of development, and evaluate what is usable in applications. Consolidation – which goes together with "robustness" – is therefore at least as relevant as innovation, to get hold of the situation of LRs (particularly important for industrial exploitation).

Moreover LRs is a 'sensitive' issue, charged of political, social, cultural, economic, and – as recently stressed – strategic implications. This makes even more crucial that, as pointed out very well in the final Euromap Report, coordination initiatives are put in place so as to avoid a two-speed situation, between languages which are interesting commercially, and today also politically, and those which are not (unfortunately the vast majority). This is in line with the spirit of LREC, that has always gone in this direction, recognising the importance and providing visibility and a large forum for discussion also to policies for LR creation in different countries and for different languages. The importance of infrastructural issues has been clearly recognised in every LREC – and also this year – as critical for a real advancement in HLT. In this respect, a quite new paradigm is emerging, in a few papers, involving initiatives aiming at open and distributed infrastructures for cooperative and controlled creation and maintenance of LRs. This is only feasible when the field as a whole has reached a level of stability and maturity. This may become the new 'vision' for LRs in the years to come.

It seems we will have many many participants to this LREC. We know we received an incredible

number of submissions. However the success has brought with it also practical and organisational concerns. We were faced with the dilemma: should we maintain the size of the last LREC and reject many submissions, or we remain faithful to the policy of providing the broadest picture of the field of LRs and evaluation, obviously preserving quality? We have decided for the second option. This meant accepting an incredibly high number of papers, between orals and posters. This decision has also forced us to decide to reduce the length of the papers to 4 pages, to avoid ending up with Proceedings of 10 or 12 volu mes! We certainly need to think about these issues for the next LREC.

At last I want to mention one desiderata, for many of us, which is slowly becoming a reality, also through LREC: by means of *less separation between Written and Spoken sessions* in this LREC, we wish to start encouraging and pushing towards more interaction and integration between the two big areas and communities. This is a must for our field to contribute, effectively and globally, to the big challenges of the 'knowledge-based society'. We have tried to introduce a novelty in how sessions are organised. In addition to the usual LREC tracks – Evaluation, Multimodality, Speech, Terminology, Written – we decided to start having some 'mixed' session, and see how well this is accepted by the participants. The goal is to favour integration among different communities. We strongly believe that integration of the two, until recently rather separate, communities is an essential step for a comprehensive approach to communication, which is made up by different modalities and their complex interactions. LREC is special also in this respect, because it is one of the few conferences that really targets all these communities at the same time and at the same level of importance.

Acknowledgments

And now it is time for thanking all those who have made this LREC possible.

First of all I deeply thank the Programme Committee (PC), a very special PC which is more a group of old friends (after so many years of working together): Joseph Mariani, President of ELRA and the wise man; Bente Maegaard, smoothing situations and always fighting for more visibility for the Evaluation area; Daniel Tapias, conveying all the expertise painfully acquired in the Las Palmas LREC; Teresa Lino, always patient and available to tackle unexpected requests, also from us; Khalid Choukri, with whom we can work almost without speaking so well we know each other, always keeping all of us 'on earth' and helping to solve unsolvable problems; and last Jan Odijk, with whom I have 'virtually' spent many weekends together (putting his patience to a severe test), trying to solve the big puzzle of this LREC program. I am sure we have made mistakes, and we ask for your forgiveness. We tried our best. I also want to say that Antonio has continued to be in many senses with us, each of us thinking so often "what he would have done?".

Then I thank with strength and sympathy the groups in Paris and Pisa, in particular: Magali Jeanmaire, with so many organisational aspects in her hands; Louis-Gabriel Pouillot, managing the web site and adapting the new Start tool; Sara Goggi, practically everything has passed through her, and she was in touch with all of you in so many occasions, as the bridge between the PC and the participants; and Sergio Rossi, unique and unequalled in finding the best practical solutions for posters and always available to help us in changing the program even twice a day. I can not list the many other tasks they carried out, but I can say for sure that without their daily work and real dedication since many months, LREC would not have been possible. Sometimes they were even organising us of the PC! Most importantly, I have seen that despite the distance (Paris -Pisa) they have worked together as a unique and wonderful team, with enthusiasm and dedication. I am sure you will notice this also during the Conference.

I thank our impressively large Scientific Committee, they did a wonderful job, almost all in time, and I am grateful to our Advisory Board, for their important cooperation. We are also indebted to the ELRA Board, and to authorities, associations, organisations, committees, agencies, companies that have supported LREC in various ways. We particularly thank Microsoft, Priberam Informática and Porto Editora for their sponsorship to the Conference. I thank the workshop organisers, and obviously all the authors, who provide the content to LREC, and give us such a broad picture of the field.

I am specially grateful to Martin Kay, Makoto Nagao and Bernard Quemada, for speaking, representing all of us, in the session in memory of Antonio.

Finally I thank the fantastic Lisbon team, headed by Teresa Lino, able to lead her group always with a smile. We have learned easily to work together, helped also by a number of PC visits to Lisbon, where we have enjoyed their nice hospitality, and through weekly trilateral phone calls (Lisbon-Paris-Pisa) to coordinate work and solve together the many big and small problems of a large conference like this. They will assist you during the days of the workshops and the conference.

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 rea lity the success of the field of LRs and Evaluation.

With all the Programme Committee, and with the Lisbon, Paris and Pisa teams, I welcome you at LREC 2004 in Lisbon, wish you a wonderful Conference, and … wait for you as numerous and enthusiastic as this time at LREC 2006.

Enjoy LREC!

<div align="right">

Nicoletta Calzolari
Chair
4th International Conference on Language Resources and Evaluation

</div>