

## LREC 2006

### 5<sup>th</sup> Conference on Language Resources and Evaluation

#### SECOND CALL FOR PAPERS



**Magazzini del Cotone Conference Center, GENOA - ITALY**

**MAIN CONFERENCE: 24-25-26 MAY 2006**

**WORKSHOPS and TUTORIALS: 22-23 and 27-28 MAY 2006**

**Conference web site: <http://www.lrec-conf.org/lrec2006>**

The fifth international conference on Language Resources and Evaluation, LREC 2006, is organised by ELRA in cooperation with a wide range of international associations and consortia, including AAMT, ACL, AFNLP, ALLC, ALTA, COCODA and Oriental COCODA, EACL, EAMT, ELSNET, ENABLER, EURALEX, Forum TAL, GWA, IAMT, ISCA, KnowledgeWeb, LDC, NEMLAR Network, SENSEVAL, SIGLEX, TEI, Technolanguage French Program, WRITE and with major national and international organisations including the European Commission – Information Society and Media Directorate General, Unit “Interfaces”.

#### **CONFERENCE AIMS**

In the Information Society, the pervasive character of Human Language Technologies (HLT) and their relevance to practically all fields of Information Society Technologies (IST) has been widely recognised. Two issues are considered particularly relevant: the availability of Language Resources (LRs) and the methods for the evaluation of resources, technologies, products and applications. Substantial mutual benefits are achieved by addressing these issues through international cooperation.

The term *language resources* refers to sets of language data and descriptions in machine readable form, such as written or spoken corpora and lexica, annotated or not, multimodal resources, grammars, terminology or domain specific databases and dictionaries, ontologies, multimedia databases, etc. LRs also cover basic software tools for their acquisition, preparation, collection, management, customisation and use. LRs are used in many types of components/systems/applications, such as software localisation and language services, language enabled information and communication services, knowledge

management, e-commerce, e-publishing, e-learning, e-government, cultural heritage, linguistic studies, etc.. This large range of usages makes the LRs infrastructure a strategic part of the e-society, where the creation of a basic set of LRs for all languages must be ensured in order to bring all languages to the same level of usability and availability.

The relevance of the evaluation for language technologies development is increasingly recognised. This involves assessing the state-of-the-art for a given technology, measuring the progress achieved within a programme, comparing different approaches to a given problem, assessing the availability of technologies for a given application, product benchmarking, and assessing system usability and user satisfaction.

The aim of the LREC conference is to provide an overview of the state-of-the-art, explore new R&D directions and emerging trends, exchange information regarding LRs and their applications, evaluation methodologies and tools, ongoing and planned activities, industrial uses and needs, requirements coming from the new e-society, both with respect to policy issues and to technological and organisational ones. LREC provides a unique forum for researchers, industrials and funding agencies from across a wide spectrum of areas to discuss problems and opportunities, find new synergies and promote initiatives for international cooperation in the areas mentioned above, in support to investigations in language sciences, progress in language technologies and development of corresponding products, services and applications.

## **CONFERENCE TOPICS**

Examples of the topics which may be addressed by papers submitted to the conference are given below.

### **Issues in the design, construction and use of Language Resources (LRs)**

- **Methodologies and tools:**
  - Guidelines, standards, specifications, models and best practices for LRs.
  - Methods, tools, procedures for the acquisition, creation, annotation, management, access, distribution, use of monolingual and multilingual LRs.
  - Methods for the extraction and acquisition of knowledge (e.g. terms, ontologies, lexical information, language modelling) from LRs, and knowledge transfer among languages.
  - Definition and requirements for a Basic and Extended LAngeage Resource Kit (BLARK, ELARK) for all languages.
  - Documentation and archiving of languages, including minority and endangered languages.
  - LRs for linguistic research in human-machine communication.
- **LRs construction & annotation:**
  - Metadata descriptions of LRs and metadata for semantic/content markup.
  - Ontologies and knowledge representation, especially with respect to HLT.
  - Terminology and NLP tools and methodologies for terminology and ontology building or mapping, term extraction, domain-specific dictionaries.
  - LRs for machine translation.
  - LRs for ubiquitous processing.
  - Availability and use of generic vs. task/domain specific LRs.
  - Multimedia and Multimodal LRs - Integration of various media and modalities in LRs (speech, vision, language).
- **LRs exploitation:**
  - Industrial production of LRs.
  - Industrial LRs requirements, user needs and community's response.
  - Exploitation of LRs in different types of applications (information extraction, information retrieval, speech dictation, translation, summarisation, web services, semantic web, semantic search, text mining, inferencing, etc.).
  - Exploitation of LRs in different types of interfaces (dialogue systems, natural language and multimodal/multisensorial interactions, etc.).

## **Issues in Human Language Technologies (HLT) evaluation**

- **Methodologies, tools and standardisation:**
  - Evaluation, validation, quality assurance of LRs.
  - Evaluation methodologies, protocols and measures.
  - Benchmarking of systems and products, resources for benchmarking and evaluation, blackbox, glassbox and diagnostic evaluation of systems.
  - From evaluation to standardisation.
  - User centred design tools and methods.
  - Evaluation of ontologies and knowledge bases by means of LR-related techniques.
  - Evaluation in written language processing: (document production and management, text retrieval, terminology extraction, message understanding, text alignment, machine translation, morphosyntactic tagging, parsing, semantic tagging, word sense disambiguation, text understanding, summarization, question answering, localization, etc.).
  - Evaluation in spoken language processing: (speech recognition and understanding, voice dictation, oral dialogue, speech synthesis, speech coding, speaker and language recognition, spoken translation, etc.).
  - Evaluation of multimedia document retrieval and search systems (including detection, indexing, filtering, alert, question answering, etc.).
  - Evaluation of multimodal systems.
- **Usability evaluation of HLT based user Interfaces:**
  - Usability and user satisfaction evaluation.
  - Psychophysical and cognitive evaluation.
  - User experience assessment.
  - Heuristic evaluation.
  - Multimodal interaction evaluation.
  - Evaluation of usability in mobile services/applications, etc.

## **General issues**

- National and international activities and projects.
- Open architectures for LRs.
- LRs and the needs/opportunities of the emerging industries.
- LRs and contributions to societal needs (e.g. e-society).
- Priorities, perspectives, strategies in national and international policies for LRs.
- Needs, possibilities, forms, initiatives of/for international cooperation, and their organisational and technological implications.
- Organisational, economical and legal issues in the construction, distribution, access and use of LRs.

## **Special Highlights**

LREC targets the integration of different types of LRs (spoken, written, and other modalities), and of the respective communities. To this end, LREC encourages submissions covering issues which are common to different types of LRs and language technologies, such as dialogue strategy, written and spoken translation, domain-specific data, multimodal communication or multimedia document processing, and will organise, in addition to the usual tracks, common sessions encompassing the different areas of LRs.

### **The 2006 Conference emphasises in particular the importance of promoting:**

- **synergies and integration between (multilingual) LRs and Semantic Web technologies,**
- **new paradigms for sharing and integrating LRs and LT coming from different sources,**
- **communication with neighbouring fields for applications in e-government and administration,**
- **common evaluation campaigns for the objective evaluation of the performances of different systems,**
- **systems and products (also industrial ones) based on large-size and high quality LRs.**

**LREC therefore encourages submissions of papers, panels, workshops, tutorials on the use of LRs in these areas.**

## **PROGRAMME**

The Scientific Programme will include invited talks, oral presentations, poster presentations and panels. **There is no difference in quality between oral presentations and poster presentations. Only the appropriateness of the type of communication (more or less interactive) to the content of the paper will be considered.**

## **ABSTRACT SUBMISSION**

On-line submission form for abstracts is now available.

You should go to the “Abstract submission” section on the LREC2006 web site (<http://www.lrec-conf.org/lrec2006/>) and follow the procedure instructions.

## **WORKSHOPS, TUTORIALS AND PANELS**

**Submission of workshop, tutorial and panel proposals** should be made by e-mail to both the following e-mail addresses:

- ✓ [lrec@ilc.cnr.it](mailto:lrec@ilc.cnr.it)
- ✓ [lrec@elda.org](mailto:lrec@elda.org)

**Proposals for workshops and tutorials** should be no longer than three pages, and include:

- **For workshops:**
  - The title
  - A brief technical description of the specific technical issues that the workshop will address
  - The reasons why the workshop is of interest
  - The names and affiliations, postal addresses, phone and fax numbers, email and web site addresses of the organising committee, which should consist of at least three people knowledgeable in the field, coming from different institutions
  - The name and the e-mail address of the member of the workshop organising committee designated as the contact person
  - The desirable duration of the workshop (half day or full day)
  - A summary of the intended call for participation
  - An estimate of the approximate audience size
  - A list of audio-visual or technical requirements and any special room requirements

The workshop proposers will be responsible for the organisational aspects (e.g. workshop call preparation and distribution, review of papers, notification of acceptance, assembling of the workshop proceedings using the ELRA specifications, etc.).

- **For tutorials:**
  - The title
  - A brief technical description of the tutorial content
  - The reasons why the tutorial is of interest
  - The names and affiliations, postal addresses, phone and fax numbers, email and web site addresses of the tutorial speakers, with brief descriptions of their technical background
  - The name and e-mail address of one tutorial speaker designated as the contact person
  - The duration of the tutorial (half day is the expected usual length)
  - An estimate of the approximate audience size
  - A list of audio-visual or technical requirements and any special room requirements

The tutorial proposers will be responsible for the organisational aspects (e.g. assembling of the tutorial material, etc.).

**Proposals for panels** should contain the following information:

- The title
- A brief technical description of the specific technical issues that the panel will address
- The reasons why the panel is of interest
- Name of the panel organiser/s; affiliation and postal address; phone and fax numbers; e-mail address; web site address
- The name and the e-mail address of the designated contact person

## **PROCEEDINGS**

The Proceedings of the conference will cover both oral and poster papers.

**The Conference package that will be provided to all will include Proceedings on CD-ROM and a printed book of Abstracts and Conference Programme. Hardcopy Proceedings will be printed on demand only.**

## **IMPORTANT DATES**

- Submission of proposals for panels, workshops and tutorials: **14 October 2005**
- Submission of proposals for oral and poster papers: **14 October 2005**
- Notification of acceptance of panels, workshops and tutorials proposals: **7 November 2005**
- Notification of acceptance of oral papers, posters: **16 January 2006**
- Final versions for the proceedings: **20 February 2006**
- Conference: **24-26 May 2006**
- Pre-conference workshops and tutorials: **22 and 23 May 2006**
- Post-conference workshops and tutorials: **27 and 28 May 2006**

Internet connections and various computer platforms and facilities will be available at the conference site. It will be possible to run non-reviewed demos of language engineering products, systems and tools. Those interested should contact the organiser of the demonstrations (details will be posted on <http://www.lrec-conf.org/lrec2006/>).

## **CONSORTIA AND PROJECT MEETINGS**

Consortia or projects wishing to take this opportunity for organising meetings should contact the ELDA office at [lrec@elda.org](mailto:lrec@elda.org).

## **CONFERENCE PROGRAMME COMMITTEE**

Nicoletta Calzolari, Istituto di Linguistica Computazionale del CNR, Pisa, Italy (Conference Chair)

Khalid Choukri, ELRA, Paris, France

Aldo Gangemi, Istituto di Scienze e Tecnologie della Cognizione del CNR, Roma, Italy

Bente Maegaard, CST, University of Copenhagen, Denmark

Joseph Mariani, LIMSI-CNRS, Orsay, France

Jan Odijk, ScanSoft, Merelbeke, Belgium and UIL-OTS, Utrecht, The Netherlands

Daniel Tapias, Telefonica Moviles, Madrid, Spain

The composition of the committees as well as instructions and addresses for registration and accommodation will be detailed on the LREC web site at <http://www.lrec-conf.org/lrec2006/>.

## ELRA

For more information about ELRA (European Language Resources Association), please contact:

Khalid Choukri, ELRA CEO  
55-57 Rue Brillat-Savarin,  
75013 Paris - France  
Tel: + 33 1 43 13 33 33  
Fax: + 33 1 43 13 33 30  
Email: [choukri@elda.org](mailto:choukri@elda.org)  
Web: <http://www.elra.info> or <http://www.elda.org/>

---

### Some facts and figures about previous LRECs

The first LREC was organised in Granada (Spain) in 1998: 197 papers and posters were presented, with about 510 registered participants from 38 different countries from all continents. Among these, the largest group came from Spain (81 participants), followed by France (75), USA (73), Germany (47), UK (43) and Italy (41). Registered participants belonged to over 325 different organisations.

LREC 2000, in Athens, had 129 oral papers and 152 posters presented, with around 600 participants from 51 different countries from all continents. Among these, the largest group came from Greece (117), followed by USA (70), France (59), Germany (45), UK (43), Japan (35) and Italy (29). Registered participants belonged to 319 different organisations.

LREC 2002, which took place in Las Palmas de Gran Canaria (Spain), attracted over 700 representatives, coming from 38 countries around the world. The following figures illustrate how successful it proved to be: for the main conference, 460 papers had been submitted and reviewed, of which 365 were presented at the conference. Most of the areas in HLT were covered (about 280 papers dealt with written resources, about 100 with spoken resources, 25 with multimodal and multimedia resources, around 50 dealt with evaluation of HLT, and 16 with terminology).

The 4th edition of the Language Resources and Evaluation Conference was held in memory of two dear friends and colleagues we lost in 2003, Angel Martin Municio and Antonio Zampolli.

LREC 2004, which took place in Lisbon (Portugal), attracted almost 1000 participants, coming from 50 countries from all the continents. Close to 800 submissions for poster and oral presentations were reviewed by the Scientific Committee: 519 were actually presented, a majority dedicated to written resources (260), 116 dealt with spoken resources, 40 with terminological issues, 57 with evaluation, 17 were on general issues, and 29 on multimodal-multimedia ones. In addition, a total of 18 satellite workshops covering various fields were organised before and after the main conference.

A new award in HLT was launched on that occasion: the ELRA Board created a prize for "Outstanding Contributions to the Advancement of Language Resources and Language Technology Evaluation", to honour the memory of its co-founder and 1<sup>st</sup> president, Antonio Zampolli. The Antonio Zampolli Prize was awarded for the first time at LREC 2004 to Fredrick Jelinek, from John Hopkins University, Baltimore, USA.

A similar number of participants is expected at LREC 2006.

**If you want to know the state-of-the-art in LT and LRs and their application in all aspects of e-society , this is the Conference to go to!**