

## THIRD CALL FOR PAPERS

Workshop on

### *Methods for the automatic acquisition of Language Resources and their evaluation methods*

*Invited Keynote Speakers*

***Kara Warburton***

Delegate at ISO TC37, LISA and Manager of Terminology at IBM

and

***Julio Gonzalo***

UNED group in Natural Language Processing and Information Retrieval

To be held in conjunction with the 7th International Language Resources  
and Evaluation Conference (LREC 2010)

**23 May 2010, Mediterranean Conference Center, Valletta, Malta**

[http://www.flarenet.eu/?q=LREC2010\\_Acquisition\\_Workshop](http://www.flarenet.eu/?q=LREC2010_Acquisition_Workshop)

*Extended Deadline for submission: 5 March 2010*

### **Tentative Workshop Agenda**

<b>9:30-9:45</b>	Welcome, Workshop Presentation
<b>9:45-10:30</b>	<b>Kara Warburton</b> "Extracting, evaluating, and preparing terminology for large-scale translation jobs".
<b>10:30-10:45</b>	Coffee Break
<b>10:45-11:30</b>	<b>Julio Gonzalo</b> "Benchmarking and Evaluation Campaigns: the good, the bad and the metrics".
<b>11:30-12:00</b>	<b>Andrejs Vasiljevs</b> "ACCURAT: Metrics for the evaluation of comparability of multilingual corpora"
<b>12:00-12:30</b>	<b>Núria Bel &amp; Valeria Quochi</b> "PANACEA: Evaluation in a factory of language resources and derivatives"
<b>12:30-13:00</b>	<b>Béatrice Daille</b> "TTC: Evaluation procedures of multilingual terminology acquired from comparable corpora"
<b>13:00-14:30</b>	Lunch
<b>14:30-16:00</b>	<b>Poster session</b>
<b>16:00-16:30</b>	Coffee Break
<b>16:30-17:30</b>	<b>Discussion session.</b> A strategy for assessing the potential and future impact of acquisition techniques: criteria for the evaluation of methods.

While methods for the Automatic Production of Language Resources are being described in scientific papers, the proposed techniques and their results are somehow unrelated because the lack of a common forum where to share experiences and to compare results.

**FlaReNet** Working Group on “Methods for the automatic construction and processing of Language Resources” in cooperation with FP7 projects: **ACCURAT** (Analysis and Evaluation of Comparable Corpora for Under Resourced Areas of Machine Translation), **PANACEA** (Platform for Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources for Human Language Technologies) and **TTC** (Terminology Extraction, Translation Tools and Comparable Corpora) organizes this LREC workshop whose aim is to be the beginning of a series of meetings that will be an enduring forum where researchers in the field can exchange information and compare results.

The two main objectives of the workshop are:

- First, to start the compilation of information about current initiatives and available and proved applications for the automatic acquisition of Language Resources and,
- Second, to start the creation of common materials for the evaluation and comparison of the results of these methods and techniques.

This first workshop is organized in a series of presentation talks and an open poster session. For the open poster session, the organizers welcome contributions related to the techniques and applications of the automatic acquisition and production of language resources and their evaluation in one of the following areas:

- Automatic extraction and pre-processing of corpus
- Alignment techniques for parallel and comparable corpus
- Bilingual dictionaries extraction, including terminological resources
- Lexical Information Acquisition
- Tree-bank automatic building
- Grammar induction
- Automatic acquisition of multimodal resources
- Automatic acquisition of sign languages

Poster papers (in the LREC format available at the conference web page and that should not be longer than 6 pages, references included) may describe past evaluation exercises as well as work in progress.

### **Keynote Speakers**

The confirmed keynote speakers are **Julio Gonzalo**, UNED group in Natural Language Processing and Information Retrieval and **Kara Warburton**, Canadian delegate at ISO TC37, Elected chair of the Terminology Special Interest group for LISA and Manager of Terminology at IBM.

**Gonzalo** is a member of the nlp.uned.es research group, where he conducts research on the application of Language Engineering to Multilingual Information Access problems and in particular in the development of evaluation metrics and methodologies. He has been involved in the coordination of CLEF (the international evaluation campaign for Multilingual Information Access applications) and WePS (Web People Search evaluation campaign).

**Warburton** is head of terminology development for IBM and is a recognized expert in terminology management. Elected chair of the LISA Terminology SIG, she is currently the project leader for the TBX submission to ISO and member of the ISO Terminology committee.

## **Important dates**

**New Deadline for Submissions: 5 march 2010**

Deadline for submission: 26 February 2010

Notification of acceptance: 18 March 2010

Final version due: 25 March 2010

Workshop: 23 May 2010

## **Submission Procedure**

Submission will be electronic using START paper submission software at the following link <https://www.softconf.com/lrec2010/FLaReNetAcquisition2010/>

When submitting a paper from the START page, authors will be asked to provide essential information about resources (in a broad sense, i.e. also technologies, standards, evaluation kits, etc.) that have been used for the work described in the paper or are a new result of your research. For further information on this new initiative, please refer to <http://www.lrec-conf.org/lrec2010/?LREC2010-Map-of-Language-Resources>."

## **Author's Kit and Template**

Authors are requested to format their manuscripts according to the guidelines specified in specified style sheet at the following link: <http://www.lrec-conf.org/lrec2010/Author-s-Kit-and-Templates.html>. Papers which do not adhere to this format will NOT be accepted for publication in the Conference Proceedings.

## **Workshop Organizing Committee**

Núria Bel	Universitat Pompeu Fabra - IULA Barcelona, Spain
Béatrice Daille	Université de Nantes Nantes, France
Andrejs Vasiljevs	Tilde - www.tilde.com Riga, Latvia

## **Confirmed Program Committee Members to date:**

Adeline Nazarenko	LIPN - The Computer Science Lab of the Paris-Nord University Paris, France
Anna Korhonen	University of Cambridge - Natural Language and Information Processing (NLIP) Group - Cambridge, United Kingdom
Anne Vilnat	LIMSI - Lab. d'Informatique Mécanique et Sciences de l'Ingénieur. Orsay Cedex, France
Eric de la Clergerie	INRIA – I. National de Recherche en Informatique et Automatique France
Gregor Thurmair	Linguatéc Munich, Germany
Montserrat Marimon	Universitat de Barcelona Barcelona, Spain
Nancy Ide	Vassar College - Department of Computer Science New York, USA
Nicoletta Calzolari	ILC-CNR, Istituto di Linguistica Computazionale Pisa, Italy
Prokopis Prokopidis	ILSP - Institute for Language Speech Processing Athens, Greece
Stelios Piperidis	ILSP - Institute for Language Speech Processing Athens, Greece
Valeria Quochi	ILC-CNR, Istituto di Linguistica Computazionale Pisa, Italy
Victoria Arranz	ELDA - Evaluations and Language resources Distribution Agency Paris, France

**Contact Person:** *Núria Bel* - [nuria.bel at upf.edu](mailto:nuria.bel@upf.edu)