

# **Collaboration: interoperability between people in the creation of language resources for less-resourced languages**

## **A SALTMIL workshop**

**May 27, 2008**

## **Workshop Programme**

- 14:30-14:35 Welcome
- 14:35-15:05 *Icelandic Language Technology Ten Years Later*  
Eiríkur Rögnvaldsson
- 15:05-15:35 *Human Language Technology Resources for Less Commonly Taught Languages: Lessons Learned Toward Creation of Basic Language Resources*  
Heather Simpson, Christopher Cieri, Kazuaki Maeda, Kathryn Baker, and Boyan Onyshkevych
- 15:35-16:05 *Building resources for African languages*  
Karel Pala, Sonja Bosch, and Christiane Fellbaum
- 16:05-16:45 COFFEE BREAK
- 16:45-17:15 *Extracting bilingual word pairs from Wikipedia*  
Francis M. Tyers and Jacques A. Pienaar
- 17:15-17:45 *Building a Basque/Spanish bilingual database for speaker verification*  
Iker Luengo, Eva Navas, Iñaki Sainz, Ibon Saratxaga, Jon Sanchez, Igor Odriozola, Juan J. Igarza and Inma Hernaez
- 17:45-18:15 *Language resources for Uralic minority languages*  
Attila Novák
- 18:15-18:45 *Eslema. Towards a Corpus for Asturian*  
Xulio Viejo, Roser Saurí and Angel Neira
- 18:45-19:00 Annual General Meeting of the SALTMIL SIG

## Workshop Organisers

**Briony Williams**

Language Technologies Unit  
Bangor University, Wales, UK

**Mikel L. Forcada**

Departament de Llenguatges i Sistemes Informàtics  
Universitat d'Alacant, Spain

**Kepa Sarasola**

Lengoaia eta Sistema Informatikoak Saila  
Euskal Herriko Unibertsitatea / University of the Basque Country

**SALTMIL**

Speech and Language Technologies for Minority Languages  
A SIG of the International Speech Communication Association  
<http://ixa2.si.ehu.es/saltmil/>

## Workshop Programme Committee

**Briony Williams**, Bangor University, Wales, UK

**Mikel L. Forcada**, Universitat d'Alacant, Spain

**Kepa Sarasola**, Euskal Herriko Unibertsitatea / University of the Basque Country

**Atelach Alemu Argaw**, Stockholm University, Sweden

**Julie Berndsen**, University College Dublin, Ireland

**Shannon Bischoff**, Universidad de Puerto Rico, Puerto Rico

**Lori Levin**, Carnegie-Mellon University, USA

**Climent Nadeu**, Universitat Politècnica de Catalunya, Spain

**Juan Antonio Pérez-Ortiz**, Universitat d'Alacant, Spain

**Bojan Petek**, University of Ljubljana, Slovenia

**Oliver Streiter**, National University of Kaohsiung, Taiwan

## Table of Contents

- 1-5 *Icelandic Language Technology Ten Years Later*  
Eiríkur Rögnvaldsson
- 7-11 *Human Language Technology Resources for Less Commonly Taught Languages: Lessons Learned Toward Creation of Basic Language Resources*  
Heather Simpson, Christopher Cieri, Kazuaki Maeda, Kathryn Baker, and Boyan Onyshkevych
- 13-18 *Building resources for African languages*  
Karel Pala, Sonja Bosch, and Christiane Fellbaum
- 19-22 *Extracting bilingual word pairs from Wikipedia*  
Francis M. Tyers and Jacques A. Pienaar
- 23-26 *Building a Basque/Spanish bilingual database for speaker verification*  
Iker Luengo, Eva Navas, Iñaki Sainz, Ibon Saratxaga, Jon Sanchez, Igor Odriozola, Juan J. Igarza and Inma Hernaez
- 27-32 *Language resources for Uralic minority languages*  
Attila Novák
- 33-38 *Eslema. Towards a Corpus for Asturian*  
Xulio Viejo, Roser Saurí and Angel Neira

## Author Index

Baker, Kathryn:	7
Bosch, Sonja:	13
Cieri, Christopher:	7
Fellbaum, Christiane:	13
Hernaez, Inma:	23
Igarza, Juanjo:	23
Luengo, Iker;	23
Maeda, Kazuaki:	7
Navas, Eva:	23
Neira, Ángel:	32
Novak, Attila:	27
Odriozola, Igor:	23
Onyshkevych, Boyan:	7
Pala, Karel:	13
Pienaar, Jacques:	19
Rögnvaldsson, Eiríkur:	1
Sainz, Iñaki:	23
Sanchez, Jon:	23
Saratxaga, Ibon:	23
Saurí, Roser:	32
Simpson, Heather:	7
Tyers, Francis:	19
Viejo, Xulio:	32