

A full-day workshop at LREC 2012

Tuesday, 22 May 2012.

Lütfi Kırdar Istanbul Exhibition and Congress Centre, Istanbul, Turkey

SALTmil: <http://ixa2.si.ehu.es/saltmil/>

AfLaT: <http://AfLaT.org/>

LREC 2012: <http://www.lrec-conf.org/lrec2012/>

WORKSHOP PROGRAMME

09:15–09:30 **Welcome / Opening Session**

09:30–10:30 **Invited Talk: Sjur Moshagen Nørstebø**. How to build language technology resources for the next 100 years

10:30–11:00 Coffee Break

11:00–13:00 **Oral papers: Resource Creation**

- **Elaine Uí Dhonnchadha, Alessio Frenda and Brian Vaughan**, Issues in Designing a Spoken Corpus of Irish.
- **Wondwossen Mulugeta and Michael Gasser**, Learning Morphological Rules for Amharic Verbs Using Inductive Logic Programming
- **Kristín Bjarnadóttir**, The Database of Modern Icelandic Inflection
- **Fadoua Ataa Allah and Siham Boulaknadel**, Natural Language Processing for Amazigh Language: Challenges and Future Directions

13:00–14:00 Lunch Break

14:00–16:00 **Oral papers: Resource Use**

- **Tommi A. Pirinen and Francis M. Tyers**. Compiling Apertium morphological dictionaries with HFST and using them in HFST applications.
- **Borbóla Siklósi, György Orosz, Attila Novák and Gábor Prószték**. Automatic structuring and correction suggestion system for Hungarian clinical records.

- **Linda Wiechete**. Constraint Grammar based Correction of Grammatical Errors for North Sàmi.
- **Michael Gasser**, Toward a Rule-Based System for English-Amharic Translation.

16:00–16:30 Coffee Break

16:30–17:30 **Poster Session**

- **Emmanuel Cartier and Paola Carrion Gonzalez**, Technological Tools for Dictionary and Corpora Building for Minority Languages: Example of the French-based Creoles.
- **Denys Duchier, Brunelle Magnana Ekoukou, Yannick Parmentier, Simon Petitjean and Emannuel Schang**, Describing Morphologically-rich Languages using Metagrammars: a Look at Verbs in Ikota.
- **Tjerk Hagemeijer, Iris Hendrickx, Abigail Tiny and Haldane Amaro**, A Corpus of Santomé.
- **Sigrún Helgadóttir, Asta Svavarsdóttir, Eiríkur Rögnvaldsson, Kristín Bjarnadóttir and Hrafn Loftsson**, The Tagged Icelandic Corpus (MM).
- **Laurette Pretorius and Sonja Bosch**, Semi-automated extraction of morphological grammars for Nguni with special reference to Southern Ndebele.
- **Björn Gambäck**, Tagging and Verifying an Amharic News Corpus.
- **Guy De Pauw, Gilles-Maurice de Schryver and Janneke van de Loo**. Resource-Light Bantu Part-of-Speech Tagging.
- **Gulshan Dovudov, Vít Suchomel and Pavel Smerk**, POS Annotated 50M Corpus of Tajik Language.

CONTEXT AND FOCUS

The 8th International Workshop of the ISCA Special Interest Group on Speech and Language Technology for Minority Languages (SALTMIL, <http://ixa2.si.ehu.es/saltmil>) and the 4th Workshop on African Language Technology (

[AfLaT2012](#)

) will be held as a joint effort in Istanbul, in May 2012, as part of the 2012 International Language Resources and Evaluation Conference (LREC 2012).

Entitled "Language technology for normalisation of less-resourced languages", the workshop is intended to continue the series of SALTMIL/LREC workshops on computational language resources for minority languages, held in Granada (1998), Athens (2000), Las Palmas de Gran Canaria (2002) and Lisbon (2004), Genoa (2006), Marrakech (2008) and Malta (2010)

and the series of AfLaT workshops, held in Athens (EACL2009), Malta (LREC2010) and Addis Ababa (AGIS11).

The Istanbul 2012 workshop aims to share information on tools and best practices, so that isolated researchers will not need to start from scratch. An important aspect will be the forming of personal contacts, which can minimize duplication of effort. There will be a balance between presentations of existing language resources, and more general presentations designed to give background information needed by all researchers.

While less-resourced languages and minority languages often struggle to find their place in a digital world dominated by only a handful of commercially interesting languages, a growing number of researchers are working on alleviating this linguistic digital divide, through localisation efforts, the development of BLARKs (basic language resource kits) and practical applications of human language technologies. The joint SALTMIL/AfLaT workshop on "Language technology for normalisation of less-resourced languages" provides a unique opportunity to connect these researchers and set up a common forum to meet and share the latest developments in the field.

ORGANIZERS (SALTMIL and AfLaT)

- * Mikel L. Forcada (SALTMIL): Machine Translation Group, School of Computing, Dublin City University, Dublin, Ireland
- * Guy De Pauw (AfLaT): CLiPS - Computational Linguistics Group, University of Antwerp, Antwerp, Belgium
- * Gilles-Maurice de Schryver (AfLaT): African Languages and Cultures, TshwaneDJe HLT, South Africa & Ghent University, Belgium
- * Kepa Sarasola (SALTMIL): Dept. of Computer Languages, University of the Basque Country
- * Francis M. Tyers (SALTMIL), Departament de Llenguatges i Sistemes Informàtics, Universitat d'Alacant, Spain
- * Peter Waiganjo Wagacha (AfLaT): School of Computing & Informatics, University of Nairobi, Nairobi, Kenya

PROGRAMME COMMITTEE

- * Iñaki Alegria: University of the Basque Country
- * Núria Bel, Universitat Pompeu Fabra, Barcelona, Spain
- * Lars Borin, Göteborgs universitet, Sweden
- * Sonja Bosch, University of South Africa, South Africa

- * Khalid Choukri (ELRA, ELDA, France)
- * Mikel L. Forcada, Universitat d'Alacant
- * Dafydd Gibbon, University of Bielefeld, Germany
- * Girish Nath Jha, Jawaharlal Nehru University, India
- * Hrafn Loftsson, Reykjavik University
- * Guy De Pauw, CLiPS, Universiteit Antwerpen
- * Laurette Pretorius, University of South Africa, South Africa
- * Lori Levin, Carnegie Mellon University, USA
- * Odetunji Odejobi, Obafemi Awolowo University, Nigeria
- * Benoît Sagot, INRIA Paris Rocquencourt & Université Paris 7, France
- * Felipe Sánchez-Martínez, Universitat d'Alacant
- * Kepa Sarasola, University of the Basque Country
- * Kevin Scannell, Saint Louis University, USA
- * Gilles-Maurice de Schryver, Universiteit Gent
- * Trond Trosterud, Universitetet i Tromsø, Norway
- * Francis M. Tyers, Universitat d'Alacant
- * Peter Waganjo Wagacha, University of Nairobi

REGISTRATION

See [Registration](#) in LREC 2012 site