

EACL 2012

**Proceedings of the Student Research Workshop at the 13th
Conference of the European Chapter of the Association for
Computational Linguistics**

26 April 2012
Avignon, France

© 2012 The Association for Computational Linguistics

ISBN 978-1-937284-19-0

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Introduction

On behalf of the Programme Committee, we are pleased to present the proceedings of the Student Research Workshop held at the 13th Conference of the European Chapter of the Association for Computational Linguistics, in Avignon, France, on April 26, 2012. Following the tradition of providing a forum for student researchers and the success of the previous workshops held in Bergen (1999), Toulouse (2001), Budapest (2003), Trento (2006) and Athens (2009), a panel of senior researchers will take part in the presentation of the papers, providing detailed comments on the work of the authors.

The Student Workshop will run as three parallel sessions, during which 10 papers will be presented. These high standard papers were carefully chosen from a total of 38 submissions coming from 20 countries, and one of them will be awarded the EACL-2012 Best Student Paper.

We would like to take this opportunity to thank the many people that have contributed in various ways to the success of the Student Workshop: the members of the Programme Committee for their evaluation of the submissions and for taking the time to provide useful detailed comments and suggestions for the improvement of papers; the nine panelists for providing detailed feedback on-site; and the students for their hard work in preparing their submissions.

We are also very grateful to the EACL for providing sponsorship for students who would otherwise be unable to attend the workshop and present their work. And finally, thanks to those who have given us advice and assistance in planning this workshop (especially Laurence Danlos, Tania Jimenez, Lluís Màrquez, Mirella Lapata and Walter Daelemans).

We hope you enjoy the Student Research Workshop.

Pierre Lison, University of Oslo
Mattias Nilsson, Uppsala University
Marta Recasens, Stanford University

EACL 2012 Student Research Workshop Co-Chairs

Program Chairs:

Pierre Lison, University of Oslo (Norway)
Mattias Nilsson, Uppsala University (Sweden)
Marta Recasens, Stanford University (USA)

Faculty Advisor:

Laurence Danlos, University Paris 7 (France)

Program Committee:

Lars Ahrenberg, Linköping University (Sweden)
Gemma Boleda, Universitat Politècnica de Catalunya (Spain)
Johan Bos, Rijksuniversiteit Groningen (Netherlands)
Claire Cardie, Cornell University (USA)
Michael Carl, Copenhagen Business School (Denmark)
Benoit Crabbé, University of Paris 7 (France)
Laurence Danlos, University of Paris 7 (France)
Koenraad de Smedt, University of Bergen (Norway)
Micha Elsner, Edinburgh University (UK)
Cédric Fairon, University of Louvain (Belgium)
Caroline Gasperin, TouchType Ltd (UK)
Nizar Habash, Columbia University (USA)
Barry Haddow, University of Edinburgh (UK)
Laura Hasler, University of Strathclyde (UK)
Graeme Hirst, University of Toronto (Canada)
Jerry Hobbs, University of Southern California (USA)
Véronique Hoste, University College Gent (Belgium)
Sophia Katrenko, Utrecht University (Netherlands)
Jun'ichi Kazama, NICT (Japan)
Dietrich Klakow, University of Saarland (Germany)
Valia Kordoni, University of Saarland (Germany)
Zornitsa Kozareva, University of Southern California (USA)
Marco Kuhlmann, Uppsala University (Sweden)
Sobha Lalitha Devi, AU-KBC Research Centre (India)
Jan Tore Lønning, University of Oslo (Norway)
M. Antònia Martí, University of Barcelona (Spain)
Haitao Mi, Chinese Academy of Sciences (China)
Marie-Francine Moens, K.U.Leuven (Belgium)
Roser Morante, University of Antwerp (Belgium)
Alessandro Moschitti, University of Trento (Italy)
Costanza Navarretta, University of Copenhagen (Denmark)
John Nerbonne, Rijksuniversiteit Groningen (Netherlands)
Constantin Orasan, University of Wolverhampton (UK)
Lilja Øvrelid, University of Oslo (Norway)
Gerald Penn, University of Toronto (Canada)
Adam Przepiórkowski, University of Warsaw (Poland)
Sujith Ravi, Yahoo Research (USA)
Jonathon Read, University of Oslo (Norway)
Horacio Rodríguez, Universitat Politècnica de Catalunya (Spain)

Dan Roth, University of Illinois at Urbana-Champaign (USA)
Marta Ruiz Costa-jussà, Barcelona Media Research Center (Spain)
David Schlangen, Bielefeld University (Germany)
Anders Søgaard, University of Copenhagen (Denmark)
Lucia Specia, University of Wolverhampton (UK)
Caroline Sporleder, University of Saarland (Germany)
Manfred Stede, University of Potsdam (Germany)
Mariona Taulé, University of Barcelona (Spain)
Stefan Thater, University of Saarland (Germany)
Antal van den Bosch, Tilburg University (Netherlands)
Erik Velldal, University of Oslo (Norway)

Panelists:

Marco Baroni, University of Trento (Italy)
Myroslava Dzikovska, University of Edinburgh (UK)
Judith Eckle-Kohler, University of Stuttgart (Germany)
Micha Elsner, University of Edinburgh (UK)
Jesús Giménez, Google (Ireland)
Graeme Hirst, University of Toronto (Canada)
Ivan Titov, Saarland University (Germany)
Marco Turchi, Joint Research Centre – IPSC (Italy)
Kalliopi Zervanou, Tilburg University (Netherlands)

Table of Contents

| | |
|---|----|
| <i>Improving Pronoun Translation for Statistical Machine Translation</i> Liane Guillou | 1 |
| <i>Cross-Lingual Genre Classification</i> Philipp Petrenz | 11 |
| <i>A Comparative Study of Reinforcement Learning Techniques on Dialogue Management</i> Alexandros Papangelis | 22 |
| <i>Manually constructed context-free grammar for Myanmar syllable structure</i> Tin Htay Hlaing | 32 |
| <i>What's in a Name? Entity Type Variation across Two Biomedical Subdomains</i> Claudiu Mihăilă and Riza Theresa Batista-Navarro | 38 |
| <i>Yet Another Language Identifier</i> Martin Majliš | 46 |
| <i>Discourse Type Clustering using POS n-gram Profiles and High-Dimensional Embeddings</i> Christelle Cocco | 55 |
| <i>Hierarchical Bayesian Language Modelling for the Linguistically Informed</i> Jan A. Botha | 64 |
| <i>Mining Co-Occurrence Matrices for SO-PMI Paradigm Word Candidates</i> Aleksander Wawer | 74 |
| <i>Improving machine translation of null subjects in Italian and Spanish</i> Lorenza Russo, Sharid Loáiciga and Asheesh Gulati | 81 |

Conference Program

Improving Pronoun Translation for Statistical Machine Translation

Liane Guillou

Cross-Lingual Genre Classification

Philipp Petrenz

A Comparative Study of Reinforcement Learning Techniques on Dialogue Management

Alexandros Papangelis

Manually constructed context-free grammar for Myanmar syllable structure

Tin Htay Hlaing

What's in a Name? Entity Type Variation across Two Biomedical Subdomains

Claudiu Mihăilă and Riza Theresa Batista-Navarro

Yet Another Language Identifier

Martin Majliš

Discourse Type Clustering using POS n-gram Profiles and High-Dimensional Embeddings

Christelle Cocco

Hierarchical Bayesian Language Modelling for the Linguistically Informed

Jan A. Botha

Mining Co-Occurrence Matrices for SO-PMI Paradigm Word Candidates

Aleksander Wawer

Improving machine translation of null subjects in Italian and Spanish

Lorenza Russo, Sharid Loáiciga and Asheesh Gulati

