

ACL 2013

**51st Annual Meeting of the
Association for Computational Linguistics**

Proceedings of the Student Research Workshop

August 5-7, 2013
Sofia, Bulgaria

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

©2013 The Association for Computational Linguistics

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

Welcome to the ACL 2013 Student Research Workshop. This workshop provides a venue for student researchers investigating topics in Computational Linguistics and Natural Language Processing to present their research, to meet potential advisors, and to receive feedback from the international research community. The workshop's goal is to aid students at multiple stages of their education: from those in the final stages of undergraduate training to those who are preparing their graduate thesis proposal.

As was the case last year, this year we solicited and accepted two categories of papers:

1. Thesis/Research Proposals: This category is appropriate for experienced students who wish to get feedback on their proposal and broader ideas for the field in order to strengthen their final research.
2. Research Papers: Most appropriate for students who are new to academic conferences. Papers in this category can describe completed work or work in progress with preliminary results.

We asked for original research by students related but not limited to the following topics:

- Cognitive modeling of language processing and psycholinguistics
- Dialogue and interactive systems
- Discourse, coreference and pragmatics
- Evaluation methods
- Information retrieval
- Language resources
- Lexical semantics and ontologies
- Low resource language processing
- Machine translation: methods, applications and evaluation
- Multilinguality in NLP
- NLP applications
- NLP and creativity
- NLP for the languages of Central and Eastern Europe and the Balkans
- NLP for the Web and social media
- Question answering
- Semantics
- Sentiment analysis, opinion mining and text classification
- Spoken language processing
- Statistical and Machine Learning methods in NLP
- Summarization and generation
- Syntax and parsing
- Tagging and chunking
- Text mining and information extraction
- Word segmentation

We selected 25 papers for publishing out of the 52 submissions (an acceptance rate of 48%) that we received from students from a wide variety of countries. Three papers were presented orally in one of the parallel sessions of the main conference. The other 22 papers were shown as posters as part of the poster session of the main conference.

We were able to provide most students with conference registration and travel stipends thanks to generous support from the U.S. National Science Foundation, the ACL Walker Student Fund, the Qatar Computing Research Institute, Google, and the EU projects META-NET and Multilingual Web LT.

The overall quality of the submissions was high and we thank our program committee for their excellent feedback and reviews. In particular, we are grateful to the members of the program committee who agreed to provide pre-submission feedback to students. We also thank our faculty advisors Steven Bethard, Preslav I. Nakov, and Feiyu Xu for their guidance and valuable feedback throughout the whole organization process of the workshop. Finally, thank you and congratulations to all of our Student Research Workshop presenters.

Student Co-Chairs

Anik Dey, *Hong Kong University of Science and Technology*
Sebastian Krause, *German Research Center for Artificial Intelligence*
Ivelina Nikolova, *Bulgarian Academy of Sciences*
Eva Maria Vecchi, *University of Trento*

Faculty Advisors

Steven Bethard, *University of Colorado at Boulder & KU Leuven*
Preslav I. Nakov, *Qatar Computing Research Institute*
Feiyu Xu, *German Research Center for Artificial Intelligence*

Program Committee Members

Eleftherios Avramidis, *German Research Center for Artificial Intelligence*
Mohit Bansal, *University of California, Berkeley*
Marco Baroni, *University of Trento*
Lee Becker, *University of Colorado Boulder*
Gemma Boleda, *University of Texas at Austin*
Benjamin Börschinger, *Macquarie University*
Svetla Boytcheva, *American University in Bulgaria*
Tommaso Caselli, *Hong Kong Polytechnic University*
Marcela Charfuelan, *German Research Center for Artificial Intelligence*
Kevin B. Cohen, *University of Colorado Denver*
Georgiana Dinu, *University of Trento*
Myroslava Dzikovsk, *University of Edinburgh*
Kuzman Ganchev, *Google*
Wei Gao, *Qatar Computing Research Institute*
Georgi Georgiev, *Ontotext AD*
Edward Grefenstette, *University of Oxford*
Kai Hong, *University of Pennsylvania*
Amjad abu Jbara, *University of Michigan*
Sandra Kübler, *Indiana University Bloomington*
Gianluca Leboni, *University of Trento & Fondazione Bruno Kessler*
Junyi Li, *University of Pennsylvania*
Hong Li, *German Research Center for Artificial Intelligence*
Wolfgang Maier, *Department for Computational Linguistics, University of Düsseldorf*
Marco Marelli, *University of Trento*
Petar Mitankin, *Sofia University*
Verginica Barbu Mititelu, *Romanian Academy Research Institute for Artificial Intelligence*
Brian Murphy, *Carnegie Mellon University*
Petya Osenova, *Bulgarian Academy of Sciences*
Alexis Palmer, *Saarland University*
Gabriella Pasi, *University of Milan Bicocca*
Vahed Qazvinian, *University of Michigan*
Markus Saers, *Hong Kong University of Science and Technology*
Ansaf Salleb-Aouissi, *Columbia University*
Lane Schwartz, *Air Force Research Laboratory*
Ang Sun, *Intelius*
John Tait, *Information Retrieval Facility*
Irina Temnikova, *Bulgarian Academy of Sciences*
Marco Turchi, *Fondazione Bruno Kessler*
Tony Veale, *University College Dublin*
Cristina Vertan, *University of Hamburg*
Alexander Volokh, *German Research Center for Artificial Intelligence*
Rui Wang, *German Research Center for Artificial Intelligence*
Michael Zock, *LIF-CNRS*

Table of Contents

<i>Categorization of Turkish News Documents with Morphological Analysis</i> Burak Kerim Akkuş and Ruket Cakici	1
<i>Crawling microblogging services to gather language-classified URLs. Workflow and case study</i> Adrien Barbaresi	9
<i>Patient Experience in Online Support Forums: Modeling Interpersonal Interactions and Medication Use</i> Annie Chen	16
<i>Detecting Metaphor by Contextual Analogy</i> Eirini Florou	23
<i>Survey on parsing three dependency representations for English</i> Angelina Ivanova, Stephan Oepen and Lilja Øvrelid	31
<i>What causes a causal relation? Detecting Causal Triggers in Biomedical Scientific Discourse</i> Claudiu Mihăilă and Sophia Ananiadou	38
<i>Text Classification based on the Latent Topics of Important Sentences extracted by the PageRank Algorithm</i> Yukari Ogura and Ichiro Kobayashi	46
<i>Automated Collocation Suggestion for Japanese Second Language Learners</i> Lis Pereira, Erlyn Manguilimotan and Yuji Matsumoto	52
<i>Understanding Verbs based on Overlapping Verbs Senses</i> Kavitha Rajan	59
<i>Topic Modeling Based Classification of Clinical Reports</i> Efsun Sarioglu, Kabir Yadav and Hyeong-Ah Choi	67
<i>Annotating named entities in clinical text by combining pre-annotation and active learning</i> Maria Skepstedt	74
<i>Multigraph Clustering for Unsupervised Coreference Resolution</i> Sebastian Martschat	81
<i>Computational considerations of comparisons and similes</i> Vlad Niculae and Victoria Yaneva	89
<i>Question Analysis for Polish Question Answering</i> Piotr Przybyła	96
<i>A Comparison of Techniques to Automatically Identify Complex Words.</i> Matthew Shardlow	103
<i>Detecting Chronic Critics Based on Sentiment Polarity and User's Behavior in Social Media</i> Sho Takase, Akiko Murakami, Miki Enoki, Naoaki Okazaki and Kentaro Inui	110
<i>Addressing Ambiguity in Unsupervised Part-of-Speech Induction with Substitute Vectors</i> Volkan Cirik	117

<i>Psycholinguistically Motivated Computational Models on the Organization and Processing of Morphologically Complex Words</i>	
Tirthankar Dasgupta	123
<i>A New Syntactic Metric for Evaluation of Machine Translation</i>	
Melania Duma, Cristina Vertan and Wolfgang Menzel	130
<i>High-quality Training Data Selection using Latent Topics for Graph-based Semi-supervised Learning</i>	
Akiko Eriguchi and Ichiro Kobayashi	136
<i>Simple, readable sub-sentences</i>	
Sigrid Klerke and Anders Søgaard	142
<i>Exploring Word Order Universals: a Probabilistic Graphical Model Approach</i>	
Xia Lu	150
<i>Robust multilingual statistical morphological generation models</i>	
Ondřej Dušek and Filip Jurčiček	158
<i>A corpus-based evaluation method for Distributional Semantic Models</i>	
Abdellah Fourtassi and Emmanuel Dupoux	165
<i>Deepfix: Statistical Post-editing of Statistical Machine Translation Using Deep Syntactic Analysis</i>	
Rudolf Rosa, David Mareček and Aleš Tamchyna	172

Conference Program

Monday August 5, 2013

(18:30 -19:45) Poster Session A

Categorization of Turkish News Documents with Morphological Analysis

Burak Kerim Akkuş and Ruket Cakici

Crawling microblogging services to gather language-classified URLs. Workflow and case study

Adrien Barbaresi

Patient Experience in Online Support Forums: Modeling Interpersonal Interactions and Medication Use

Annie Chen

Detecting Metaphor by Contextual Analogy

Eirini Florou

Survey on parsing three dependency representations for English

Angelina Ivanova, Stephan Oepen and Lilja Øvrelid

What causes a causal relation? Detecting Causal Triggers in Biomedical Scientific Discourse

Claudiu Mihăilă and Sophia Ananiadou

Text Classification based on the Latent Topics of Important Sentences extracted by the PageRank Algorithm

Yukari Ogura and Ichiro Kobayashi

Automated Collocation Suggestion for Japanese Second Language Learners

Lis Pereira, Erlyn Manguilimotan and Yuji Matsumoto

Understanding Verbs based on Overlapping Verbs Senses

Kavitha Rajan

Topic Modeling Based Classification of Clinical Reports

Efsun Sarioglu, Kabir Yadav and Hyeong-Ah Choi

Annotating named entities in clinical text by combining pre-annotation and active learning

Maria Skeppstedt

Monday August 5, 2013 (continued)

(19:45 – 21:00) Poster Session B

Multigraph Clustering for Unsupervised Coreference Resolution

Sebastian Martschat

Computational considerations of comparisons and similes

Vlad Niculae and Victoria Yaneva

Question Analysis for Polish Question Answering

Piotr Przybyła

A Comparison of Techniques to Automatically Identify Complex Words.

Matthew Shardlow

Detecting Chronic Critics Based on Sentiment Polarity and User's Behavior in Social Media

Sho Takase, Akiko Murakami, Miki Enoki, Naoaki Okazaki and Kentaro Inui

Addressing Ambiguity in Unsupervised Part-of-Speech Induction with Substitute Vectors

Volkan Cirik

Psycholinguistically Motivated Computational Models on the Organization and Processing of Morphologically Complex Words

Tirthankar Dasgupta

A New Syntactic Metric for Evaluation of Machine Translation

Melania Duma, Cristina Vertan and Wolfgang Menzel

High-quality Training Data Selection using Latent Topics for Graph-based Semi-supervised Learning

Akiko Eriguchi and Ichiro Kobayashi

Simple, readable sub-sentences

Sigrid Klerke and Anders Søgaard

Exploring Word Order Universals: a Probabilistic Graphical Model Approach

Xia Lu

Tuesday August 6, 2013

Oral Presentation

- 16:45 *Robust multilingual statistical morphological generation models*
Ondřej Dušek and Filip Jurčiček
- 17:05 *A corpus-based evaluation method for Distributional Semantic Models*
Abdellah Fourtassi and Emmanuel Dupoux
- 17:25 *Deepfix: Statistical Post-editing of Statistical Machine Translation Using Deep Syntactic Analysis*
Rudolf Rosa, David Mareček and Aleš Tamchyna

