

EACL 2012

**13th Conference of the European Chapter of the  
Association for Computational Linguistics**

**Proceedings of ROBUS-UNSUP 2012: Joint Workshop on  
Unsupervised and Semi-Supervised Learning in NLP**

April 23 - 27 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](mailto:acl@aclweb.org)

## Foreword

In recent years, there has been an increased interest in minimizing the need for annotated data in NLP. Significant progress has been made in the development of both semi-supervised and unsupervised learning approaches. Semi-supervised approaches are already showing remarkable empirical success, with models that exploit mixtures of labeled and unlabeled data obtaining best results in several tasks. Although unsupervised approaches have proved more challenging than semi-supervised ones, their further development is particularly important because they carry the highest potential in terms of avoiding the annotation cost. Such approaches can be applied to any language or genre for which adequate raw text resources are available.

This workshop aimed to bring together researchers dedicated to designing and evaluating unsupervised and semi-supervised learning algorithms for NLP problems. The workshop accepted submissions in any topic related to unsupervised and semi-supervised learning. However, specific focus was given to two special themes: robust algorithms and explorations of the continuum from unsupervised to semi-supervised learning.

**Robust Algorithms:** By more robust unsupervised or semi-supervised learning algorithms we mean algorithms with few parameters that give good results across different data sets and/or different applications. Many algorithms including EM, self-training and co-training are very parameter-sensitive, and parameter tuning has therefore become an important research topic. We explicitly encourage submissions that present robust algorithms or evaluate the robustness of known algorithms.

**The Continuum from Unsupervised to Semi-Supervised Learning:** The distinction between unsupervised and semi-supervised learning approaches is often not very clear, and we explicitly encourage submissions about grey-zone approaches such as weak and indirect supervision, learning from nearly free annotations (e.g. html mark-up), joint learning from several modalities, cross-language adaptation, and learning with knowledge-based priors or posteriors.

The workshop was carried out as a joint workshop between two workshop series. Predecessors are UNSUP-2011 — First Workshop on Unsupervised Learning in NLP (held at EMNLP 2011, Edinburgh, Scotland, UK) and ROBUS 2011 - Workshop on Robust Unsupervised and Semisupervised Methods in Natural Language Processing (in conjunction with RANL 2011, Hissar, Bulgaria). We invited technical papers as well as survey and position papers.

For the workshop, we received 10 submissions, of which we accepted 7.

Avignon, April 2012

The organizing committee



**Organizers:**

Omri Abend (Hebrew University of Jerusalem)  
Chris Biemann (TU Darmstadt)  
Anna Korhonen (University of Cambridge)  
Ari Rappoport (Hebrew University of Jerusalem)  
Roi Reichart (MIT)  
Anders Søgaard (University of Copenhagen)

**Program Committee:**

Steven Abney (University of Michigan, USA)  
Jason Baldridge (University of Texas at Austin, USA)  
Phil Blunsom (Oxford University, UK)  
Stefan Bordag (ExB Research & Development)  
Sam Brody (Rutgers University, USA)  
Alexander Clark (Royal Holloway, University of London, UK)  
Shay Cohen (Columbia University, USA)  
Trevor Cohn (University of Sheffield, UK)  
Gregory Druck (University of Massachusetts Amherst, USA)  
Eugenie Giesbrecht (FZI Karlsruhe)  
Joao Graca (University of Pennsylvania, USA)  
Florian Holz (University of Leipzig)  
Jonas Kuhn (University of Stuttgart)  
Percy Liang (University of Stanford, USA)  
Suresh Manandhar (University of York, UK)  
Diana McCarthy (Lexical Computing, Ltd., UK)  
Preslav Nakov (Qatar Computing Research Institute, Qatar Foundation)  
Roberto Navigli (University of Rome, Italy)  
Vincent Ng (UT Dallas, USA)  
Andreas Vlachos (University of Cambridge, UK)  
Reinhard Rapp (JG University of Mainz)  
Andrew Rosenberg (CUNY, USA)  
Sabine Schulte im Walde (University of Stuttgart, Germany)  
Noah A. Smith (CMU, USA)  
Valentin I. Spitzkovsky (University of Stanford, USA)  
Torsten Zesch (TU Darmstadt)



## Table of Contents

<i>Fast Unsupervised Dependency Parsing with Arc-Standard Transitions</i> Mohammad Sadegh Rasooli and Hesham Faili .....	1
<i>Dependency-Based Open Information Extraction</i> Pablo Gamallo, Marcos Garcia and Santiago Fernández-Lanza .....	10
<i>Sweeping through the Topic Space: Bad luck? Roll again!</i> Martin Riedl and Chris Biemann .....	19
<i>Clustered Word Classes for Preordering in Statistical Machine Translation</i> Sara Stymne .....	28
<i>Improving Distantly Supervised Extraction of Drug-Drug and Protein-Protein Interactions</i> Tamara Bobic, Roman Klinger, Philippe Thomas and Martin Hofmann-Apitius .....	35
<i>Robust Induction of Parts-of-Speech in Child-Directed Language by Co-Clustering of Words and Contexts</i> Richard E Leibbrandt and David MW Powers .....	44
<i>Dependency Parsing Domain Adaptation using Transductive SVM</i> Antonio Valerio Miceli Barone and Giuseppe Attardi .....	55





## Conference Program

*Fast Unsupervised Dependency Parsing with Arc-Standard Transitions*

Mohammad Sadegh Rasooli and Heshaam Faili

*Dependency-Based Open Information Extraction*

Pablo Gamallo, Marcos Garcia and Santiago Fernández-Lanza

*Sweeping through the Topic Space: Bad luck? Roll again!*

Martin Riedl and Chris Biemann

*Clustered Word Classes for Preordering in Statistical Machine Translation*

Sara Stymne

*Improving Distantly Supervised Extraction of Drug-Drug and Protein-Protein Interactions*

Tamara Bobic, Roman Klinger, Philippe Thomas and Martin Hofmann-Apitius

*Robust Induction of Parts-of-Speech in Child-Directed Language by Co-Clustering of Words and Contexts*

Richard E Leibbrandt and David MW Powers

*Dependency Parsing Domain Adaptation using Transductive SVM*

Antonio Valerio Miceli Barone and Giuseppe Attardi

