NAACL HLT 2013

The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Proceedings of the Main Conference

9–14 June 2013
Westin Peachtree Plaza Hotel
Atlanta, Georgia

Sponsors

NAACL HLT 2013 gratefully acknowledges the following sponsors for their support.

Gold Level

Bronze Level

Student Best Paper Award and Student Lunch Sponsor

Student Volunteer Sponsor

Conference Bags

©2013 The Association for Computational Linguistics

209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-47-3

General Chair Preface

Welcome everyone!

It is my pleasure to welcome you all to Atlanta, Georgia, for the 2013 NAACL Human Language Technologies conference. This is a great opportunity to reconnect with old friends and make new acquaintances, learn the latest in your own field and become curious about new areas, and also to experience Atlanta's warm southern hospitality. That hospitality starts with Priscilla Rasmussen! Priscilla thinks about everything that we all take for granted: the registration that just took place, the rooms in which we sit, the refreshments that keep us energized, and the social events that make this conference so fun, and many other details that you would miss if they weren't there. Please introduce yourself and say hi. Priscilla is the backbone of the NAACL organization. Thank you!

This conference started a year ago, when Hal Daumé III and Katrin Kirchhoff graciously agreed to be program co-chairs. It is no exaggeration to say how much their dedication has shaped this conference and how grateful I am for their initiative and hard work. Thank you Hal and Katrin, especially for all the fun discussion that made the work light and the year go by fast! This conference could not have happened with you.

Thanks go to the entire organizing committee. As I am writing this to be included in the proceedings, I am grateful for the fantastic detailed and proactive work by Colin Cherry and Matt Post, the publications chairs. The tutorials chairs, Katrin Erk and Jimmy Lin, selected, and solicited, 6 tutorials to present in depth material on some of the diverse topics represented in our community. Chris Dyer and Derrick Higgins considered which projects shine best when shown as a demonstration. The workshops chairs for NAACL, Sujith Ravi and Luke Zettlemoyer, worked jointly with ACL and EMNLP to select the workshops to be held at NAACL. They also worked with ICML 2013 to co-host workshops that bridge the two communities, in addition to the Joint NAACL/ICML symposium.

Posters from the student research workshop are part of the poster and demonstrations session on Monday night. This is a great opportunity for the students to be recognized in the community and to benefit from lively discussion of their presentations (attendees take note!) Annie Louis and Richard Socher are the student research workshop chairs, and Julia Hockenmaier and Eric Ringger generously share their wisdom as the faculty advisors. The student research workshop itself will be held on the first day of workshops. There are so many people who contribute their time to the behind-the-scenes organization of the conference, without which the conference cannot take place. Asking for money is probably not a natural fit for anyone, but Chris Brew worked on local sponsorship, and Dan Bikel and Patrick Pantel worked to obtain sponsorship across the ACL conferences this year - thank you! Jacob Eisenstein had the more fun role of distributing money as the student volunteer coordinator, and we thank all of the student volunteers who will be helping to run a smooth conference. Kristy Boyer kept the communication "short and tweet" using a variety of social media (and old-fashioned media too). An important part of the behind-the-scenes efforts that enable a conference like NAACL to come together are the sponsors. We thank all of the sponsors for the contributions to the conference , both for the general funding made available as well as the specific programs that are funded through sponsorship. You can read more about these sponsors in our conference handbook.

This year there are several initiatives, and if successful, we hope they'll be part of NAACL conferences

in the future. One is to make the proceedings available prior to the conference; we hope you will benefit from the extra time to read the papers beforehand. Another is for tutorials and all oral presentations to be recorded on video and made available post-conference. We are also delighted to host presentations, in both oral and poster formats, from the new Transactions of the ACL journal, to enhance the impact these will already have as journal publications. Finally, Matt Post is creating a new digital form of conference handbook to go with our digital age; thanks also go to Alex Clemmer who has prepared the paper copy that you may be reading right now. We hope you use the #NAACL2013 tag when you are tweeting about the conference or papers at the conference; together, we'll be creating a new social media corpus to explore.

Once again, we are pleased to be co-located with *SEM conference, and the SemEval workshop. We are lucky to have ICML 2013 organized so close in time and place. Several researchers who span the two communities have reconvened the Joint NAACL/ICML symposium on June 15, 2013. In addition, two workshops that address areas of interest to both NAACL and ICML members have been organized on June 16th, as part of the ICML conference.

NAACL 2013 has given me a great appreciation for the volunteering that is part of our culture. Besides the organizing committee itself, we are guided by the NAACL executive board, who think about questions with a multi-year perspective. I also want to recognize the members who first initiated and now maintain the ACL Anthology, where all of our published work will be available to all in perpetuity, a fabulous contribution and one that distinguishes our academic community.

Have a fun conference!

Lucy Vanderwende, Microsoft Research
NAACL HLT 2013 General Chair

Program Chair Preface

Welcome to NAACL HLT 2013 in Atlanta, Georgia. We have an exciting program consisting of six tutorials, 24 sessions of talks (both for long and short papers), an insane poster madness session that includes posters from the newly revamped student research workshop, ten workshops and two additional cross-pollination workshops held jointly with ICML (occurring immediately after NAACL HLT, just one block away). There are a few innovations in the conference this year, the most noticeable of which is the twitter channel #naacl2013 and the fact that we are the first conference to host papers published in the Transactions of the ACL journal – there are six such papers in our program, marked as [TACL]. We are very excited about our two invited talks, one on Monday morning and one Wednesday morning. The first is by Gina Kuperberg, who will talk about "Predicting Meaning: What the Brain tells us about the Architecture of Language Comprehension." The second presenter is our own Kathleen KcKeown, who will talk about "Natural Language Applications from Fact to Fiction."

The morning session on Tuesday includes the presentation of best paper awards to two worthy recipients. The award for Best Short Paper goes to Marta Recasens, Marie-Catherine de Marneffe and Christopher Potts for their paper "The Life and Death of Discourse Entities: Identifying Singleton Mentions" The award for Best Student Paper goes to the long paper "Automatic Generation of English Respellings" by Bradley Hauer and Greg Kondrak. We gratefully acknowledge IBM's support for the Student Best Paper Award. Finally, many thanks to the Best Paper Committee for selecting these excellent papers!

The complete program includes 95 long papers (of which six represent presentations from the journal Transactions of the ACL, a first for any ACL conference!) and 51 short papers. We are excited that the conference is able to present such a dynamic array of papers, and would like to thank the authors for their great work. We worked hard to keep the conference to three parallel sessions at any one time to hopefully maximize a participant's ability to see everything she wants! This represents an acceptance rate of 30% for long papers and 37% for short papers. More details about the distribution across areas and other statistics will be made available in the NAACL HLT Program Chair report on the ACL wiki: `http://aclweb.org/adminwiki/index.php?title=Reports`

The review process for the conference was double-blind, and included an author response period for clarifying reviewers' questions. We were very pleased to have the assistance of 350 reviewers, each of whom reviewed an average of 3.7 papers, in deciding the program. We are especially thankful for the reviewers who spent time reading the author responses and engaging other reviewers in the discussion board. Assigning reviewers would not have been possible without the hard work of Mark Dredze and his miracle assignment scripts. Furthermore, constructing the program would not have been possible without 22 excellent area chairs forming the Senior Program Committee: Eugene Agichtein, Srinivas Bangalore, David Bean, Phil Blunsom, Jordan Boyd-Graber, Marine Carpuat, Joyce Chai, Vera Demberg, Bill Dolan, Doug Downey, Mark Dredze, Markus Dreyer, Sanda Harabagiu, James Henderson, Guy Lapalme, Alon Lavie, Percy Liang, Johanna Moore, Ani Nenkova, Joakim Nivre, Bo Pang, Zak Shafran, David Traum, Peter Turney, and Theresa Wilson. Area chairs were responsible for managing paper assignments, collating reviewer responses, handling papers for other area chairs or program chairs who had conflicts of interest, making recommendations for paper acceptance or rejection, and nominating best papers from their areas. We are very grateful for the time and energy

that they have put into the program.

There are a number of other people that we interacted with who deserve a hearty thanks for the success of the program. Rich Gerber and the START team at Softconf have been invaluable for helping us with the mechanics of the reviewing process. Matt Post and Colin Cherry, as publications co-chairs, have been very helpful in assembling the final program and coordinating the publications of the workshop proceedings. There are several crucial parts of the overall program that were the responsibility of various contributors, including Annie Louis, Richard Socher, Julia Hockenmaier and Eric Ringger (Student Research Workshop chairs, who did an amazing job revamping the SRW); Jimmy Lin and Katrin Erk (Tutorial Chairs); Luke Zettlemoyer and Sujith Ravi (Workshop Chairs); Chris Dyer and Derrick Higgins (Demo Chairs); Jacob Eisenstein (Student Volunteer Coordinator); Chris Brew (Local Sponsorship Chair); Patrick Pantel and Dan Bikel (Sponsorship Chairs); and the new-founded Publicity chair who handled #naacl2013 tweeting among other things, Kristy Boyer.

We would also like to thank Chris Callison-Burch and the NAACL Executive Board for guidance during the process. Michael Collins was amazingly helpful in getting the inaugural TACL papers into the NAACL HLT conference. Priscilla Rasmussen deserves, as always, special mention and warmest thanks as the local arrangements chair and general business manager. Priscilla is amazing and everyone who sees her at the conference should thank her.

Finally, we would like to thank our General Chair, Lucy Vanderwende, for both her trust and guidance during this process. She helped turn the less-than-wonderful parts of this job to roses, and her ability to organize an incredibly complex event is awe inspiring. None of this would have happened without her.

We hope that you enjoy the conference!

Hal Daumé III, University of Maryland
Katrin Kirchhoff, University of Washington

Organizing Committee

General Conference Chair

Lucy Vanderwende, Microsoft Research

Program Committee Chairs

Hal Daumé III, University of Maryland
Katrin Kirchhoff, University of Washington

Local Arrangements

Priscilla Rassmussen

Workshop Chairs

Luke Zettlemoyer, University of Washington
Sujith Ravi, Google

Tutorial Chairs

Jimmy Lin, University of Maryland
Katrin Erk, University of Texas at Austin

Student Research Workshop

Chairs:
Annie Louis, University of Pennsylvania
Richard Socher, Stanford University
Faculty Advisors:
Julia Hockenmaier, University of Illinois at Urbana-Champaign
Eric Ringger, Brigham Young University

Student Volunteer Coordinator

Jacob Eisenstein, School of Interactive Computing, Georgia Tech

Demonstrations Chairs

Chris Dyer, Carnegie Mellon University
Derrick Higgins, Educational Testing Service

Local Sponsorship Chair

Chris Brew, Educational Testing Service

NAACL Sponsorship Chairs

Patrick Pantel, Microsoft Research
Dan Bikel, Google

Publications Chairs

Matt Post, Johns Hopkins University
Colin Cherry, National Research Council, Canada

Publicity Chair

Kristy Boyer, North Carolina State University

Program Committee

Program Committee Chairs

Hal Daumé III, University of Maryland
Katrin Kirchhoff, University of Washington

Area Chairs

Phonology and Morphology, Word Segmentation
Markus Dreyer (SDL Language Weaver)
Syntax, Tagging, Chunking and Parsing
Joakim Nivre (Uppsala University)
James Henderson (Université de Genève)
Semantics
Percy Liang (Stanford University)
Peter Turney (National Research Council of Canada)
Multimodal NLP
Srinivas Bangalore (AT&T)
Discourse, Dialogue, Pragmatics
David Traum (Institute for Creative Technologies)
Joyce Chai (Michigan State University)
Linguistic Aspects of CL
Vera Demberg (Saarland University)
Summarization
Guy Lapalme (Université de Montréal)
Generation
Johanna Moore (University of Edinburgh)
ML for Language Processing
Phil Blunsom (University of Oxford)
Mark Dredze (Johns Hopkins University)

Machine Translation
Alon Lavie (Carnegie Mellon University)
Marine Carpuat (National Research Council of Canada)
Information Retrieval and QA
Eugene Agichtein (Emory University)
Information Extraction
Doug Downey (Northwestern University)
Sanda Harabagiu (University of Texas at Dallas)
Spoken Language Processing
Zak Shafran (Oregon Health and Science University)
Sentiment Analysis and Opinion Mining
Bo Pang (Cornell University)
Theresa Wilson (Johns Hopkins University)
NLP-enabled Technology
David Bean (TDW)
Document Categorization and Topic Clustering
Jordan Boyd-Graber (University of Maryland)
Social Media Analysis and Processing
Bill Dolan (Microsoft Research)
Language Resources and Evaluation Methods
Ani Nenkova (University of Pennsylvania)

Primary Reviewers

Ahmed Abbasi
Mikhail Ageev
Eneko Agirre
Gregory Aist
Jan Alexandersson
Nicholas Andrews
David Andrzejewski
Gabor Angeli
Yoav Artzi
Michael Auli
Michiel Bacchiani
Anton Bakalov
Kirk Baker
Tyler Baldwin
Marco Baroni
Roberto Basili
Beata Beigman Klebanov
Kedar Bellare
Patrice Bellot
Emily M. Bender
Jonathan Berant
Justin Betteridge
Chandra Bhagavatula
Arianna Bisazza
Nathan Bodenstab
Danushka Bollegala
Alexandre Bouchard
Jordan Boyd-Graber
S.R.K. Branavan
Thorsten Brants
Chris Brew
Wray Buntine
David Burkett
Jill Burstein
Aoife Cahill
Chris Callison-Burch
Nicoletta Calzolari
Nicola Cancedda
Sandra Carberry
Claire Cardie
Xavier Carreras
Daniel Cer
Nate Chambers
Ming-Wei Chang
Boxing Chen
Chen Chen
David Chen
Colin Cherry
David Chiang
Yejin Choi
Jennifer Chu-Carroll
Stephen Clark
James Clarke
Martin Cmejrek
Shay Cohen
Trevor Cohn
Kevyn Collins-Thompson
John Conroy
Aron Culotta
James Cussens
Lyne Da Sylva
Ido Dagan
Robert Daland
Bhavana Dalvi
William Darling
Dipanjan Das

Pradipto Das
Eric De La Clergerie
Steve DeNeefe
John DeNero
David DeVault
Michael Denkowski
Jacob Devlin
Laura Dietz
Gregory Druck
Lan Du
Chris Dyer
Koji Eguchi
Vladimir Eidelman
Jacob Eisenstein
Jason Eisner
Ahmad Emami
Andrea Esuli
Anthony Fader
Atefeh Farzindar
Anna Feldman
Radu Florian
George Foster
Jennifer Foster
Mary Ellen Foster
Bob Frank
Dayne Freitag
Michel Galley
Michael Gamon
Sudeep Gandhe
Kavita Ganesan
Claire Gardent
Matt Gardner
Niyu Ge
Matthew Gerber
George Giannakopoulos
Daniel Gildea
Daniel Gillick
Kevin Gimpel
Filip Ginter
Yoav Goldberg
Dan Goldwasser
Sharon Goldwater
Dave Golland
Kyle Gorman
Cyril Goutte
Amit Goyal
Joao Graca
Brigitte Grau
Edward Grefenstette
Justin Grimmer
Carlos Gómez-Rodríguez
Nizar Habash
Barry Haddow
Eva Hajicova
John Hale
David Hall
Keith Hall
Greg Hanneman
Claudia Hauff
Xiaodong He
Kenneth Heafield
James Henderson
John Henderson
Ulf Hermjakob
Derrick Higgins
Graeme Hirst
Anna Hjalmarsson
Hieu Hoang
Julia Hockenmaier
Matthew Hoffman
Kristy Hollingshead
Yuening Hu
Fei Huang
Liang Huang
Minlie Huang
Zhongqiang Huang
Rebecca Hwa
Diana Inkpen
Ann Irvine
Abe Ittycheriah
Jagadeesh Jagarlamudi
Jiarong Jiang
Howard Johnson
Michael Johnston
David Jurgens
Alexander Kain
Pallika Kanani
Anna Kazantseva
Alistair Kennedy
Tracy Holloway King
Brian Kingsbury
Alexandre Klementiev
Philipp Koehn
Rob Koeling
Moshe Koppel
Alexander Kotov
Jayant Krishnamurthy
Sandra Kuebler
Marco Kuhlmann
Jonas Kuhn
Roland Kuhn
Seth Kulick
Shankar Kumar
Oren Kurland
Tom Kwiatkowski
Yoong Keok Lee
Maider Lehr
Alessandro Lenci
Gregor Leusch
Rivka Levitan
Fangtao Li
Mu Li
Shoushan Li
Percy Liang
Jimmy Lin
Xiao Ling
Diane Litman
Ding Liu
Qun Liu
Yang Liu
Adam Lopez
Annie Louis
Xiaofei Lu
Yue Lu
Michael Lucas
Xiaoqiang Luo
Klaus Macherey
Wolfgang Macherey
Nitin Madnani
Suresh Manandhar
Gideon Mann
Lluis Marquez
Erwin Marsi
Andre Martins
Yuval Marton

Sameer Maskey
Yuji Matsumoto
Evgeny Matusov
Arne Mauser
Diana McCarthy
David McClosky
Arul Menezes
Florian Metze
Donald Metzler
Haitao Mi
Rada Mihalcea
Minel Minel
Margaret Mitchell
Yusuke Miyao
Saif Mohammad
Taesun Moon
Robert Moore
Roser Morante
Louis-Philippe Morency
Preslav Nakov
Nava Nava
Roberto Navigli
Mark-Jan Nederhof
Hwee Tou Ng
Vincent Ng
Patrick Nguyen
Viet-An Nguyen
Joakim Nivre
Brendan O'Connor
Stephan Oepen
Miles Osborne
Myle Ott
Karolina Owczarzak
Martha Palmer
Sinno J. Pan
Bo Pang
Rebecca J. Passonneau
Siddharth Patwardhan
Michael Paul
Lisa Pearl
Ted Pedersen
Gerald Penn
Slav Petrov
Thierry Poibeau
Heather Pon-Barry
Hoifung Poon
Andrei Popescu-Belis
Matthew Purver
Chris Quirk
Reinhard Rapp
Roi Reichart
Ehud Reiter
Jason Riesa
Stefan Riezler
Ellen Riloff
Eric Ringger
Alan Ritter
Brian Roark
Antonio Roque
Carolyn Rose
Andrew Rosenberg
Markus Saers
Alicia Sagae
Kenji Sagae
Horacio Saggion
Saurav Sahay
Mark Sammons
Murat Saraclar
Anoop Sarkar
Giorgio Satta
Roser Saurí
Asad Sayeed
David Schlangen
Judith Schlesinger
Lane Schwartz
Holger Schwenk
Hendra Setiawan
Zak Shafran
Libin Shen
Wade Shen
Michel Simard
Sameer Singh
Jason Smith
Nathaniel Smith
Noah A. Smith
Swapna Somasundaran
Lucia Specia
Valentin Spitkovsky
Caroline Sporleder
Vivek Srikumar
Keith Stevens
Mark Stevenson
Matthew Stone
Veselin Stoyanov
Fabian Suchanek
Ang Sun
Mihai Surdeanu
Jun Suzuki
Stan Szpakowicz
Partha Talukdar
Christoph Tillmann
Ivan Titov
Kristina Toutanova
Reut Tsarfaty
Oren Tsur
Benjamin Van Durme
Josef van Genabith
Vincent Vanhoucke
Enrique Vidal
Karthik Visweswariah
Adam Vogel
Stephan Vogel
Xiaojun Wan
Haifeng Wang
Taro Watanabe
Bonnie Webber
David Weir
Michael White
Jan Wiebe
Shuly Wintner
Kristian Woodsend
Bing Xiang
Peng Xu
Hui Yang
Muyun Yang
Yi Yang
Tae Yano
Limin Yao
Mahsa Yarmohammadi
Alexander Yates
Wen-tau Yih
Yisong Yue
Rabih Zbib
Richard Zens
Luke Zettlemoyer

Ke Zhai
Congle Zhang
Lei Zhang
Min Zhang
Shiqi Zhao
Tiejun Zhao
Bowen Zhou
Jun Zhu
Xiaodan Zhu
Chengqing Zong
Geoffrey Zweig

Secondary Reviewers

JH
Karteek Addanki
Neil Ashton
Daniel Blanchard
Hailong Cao
Dave Carter
Glen Coppersmith
Daniel Dahlmeier
David Etter
Paul Felt
Francisco Guzman
Robbie Haertel
Khairun Nisa Hassanali
Kriste Krstovski
Jun Lang
Wang Ling
Hito Matsushita
Hans Moen
Kevin Seppi
Jun Sun
Xavier Tannier
Svitlana Volkova
Haochang Wang
Xinglong Wang
Mo Yu
Feifei Zhai
Bo Zhao
Kai Zhao
Xiaoning Zhu

Table of Contents

Model With Minimal Translation Units, But Decode With Phrases
Nadir Durrani, Alexander Fraser and Helmut Schmid 1

Beyond Left-to-Right: Multiple Decomposition Structures for SMT
Hui Zhang, Kristina Toutanova, Chris Quirk and Jianfeng Gao 12

Improved Reordering for Phrase-Based Translation using Sparse Features
Colin Cherry 22

Simultaneous Word-Morpheme Alignment for Statistical Machine Translation
Elif Eyigöz, Daniel Gildea and Kemal Oflazer 32

Multi-faceted Event Recognition with Bootstrapped Dictionaries
Ruihong Huang and Ellen Riloff 41

Named Entity Recognition with Bilingual Constraints
Wanxiang Che, Mengqiu Wang, Christopher D. Manning and Ting Liu 52

Minimally Supervised Method for Multilingual Paraphrase Extraction from Definition Sentences on the Web
Yulan Yan, Chikara Hashimoto, Kentaro Torisawa, Takao Kawai, Jun'ichi Kazama and Stijn De Saeger 63

Relation Extraction with Matrix Factorization and Universal Schemas
Sebastian Riedel, Limin Yao, Andrew McCallum and Benjamin M. Marlin 74

Extracting the Native Language Signal for Second Language Acquisition
Ben Swanson and Eugene Charniak 85

An Analysis of Frequency- and Memory-Based Processing Costs
Marten van Schijndel and William Schuler 95

Cross-Lingual Semantic Similarity of Words as the Similarity of Their Semantic Word Responses
Ivan Vulić and Marie-Francine Moens 106

Combining multiple information types in Bayesian word segmentation
Gabriel Doyle and Roger Levy 117

Training Parsers on Incompatible Treebanks
Richard Johansson 127

Learning a Part-of-Speech Tagger from Two Hours of Annotation
Dan Garrette and Jason Baldridge 138

Experiments with Spectral Learning of Latent-Variable PCFGs
Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster and Lyle Ungar 148

Representing Topics Using Images
Nikolaos Aletras and Mark Stevenson 158

Drug Extraction from the Web: Summarizing Drug Experiences with Multi-Dimensional Topic Models
Michael J. Paul and Mark Dredze 168

Towards Topic Labeling with Phrase Entailment and Aggregation
Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng and Shafiq Joty 179

Topic Segmentation with a Structured Topic Model
Lan Du, Wray Buntine and Mark Johnson 190

Text Alignment for Real-Time Crowd Captioning
Iftekhar Naim, Daniel Gildea, Walter Lasecki and Jeffrey P. Bigham 201

Discriminative Joint Modeling of Lexical Variation and Acoustic Confusion for Automated Narrative Retelling Assessment
Maider Lehr, Izhak Shafran, Emily Prud'hommeaux and Brian Roark 211

Using Out-of-Domain Data for Lexical Addressee Detection in Human-Human-Computer Dialog
Heeyoung Lee, Andreas Stolcke and Elizabeth Shriberg 221

Segmentation Strategies for Streaming Speech Translation
Vivek Kumar Rangarajan Sridhar, John Chen, Srinivas Bangalore, Andrej Ljolje and Rathinavelu Chengalvarayan 230

Enforcing Subcategorization Constraints in a Parser Using Sub-parses Recombining
Seyed Abolghasem Mirroshandel, Alexis Nasr and Benoît Sagot 239

Large-Scale Discriminative Training for Statistical Machine Translation Using Held-Out Line Search
Jeffrey Flanigan, Chris Dyer and Jaime Carbonell 248

Measuring Term Informativeness in Context
Zhaohui Wu and C. Lee Giles 259

Unsupervised Learning Summarization Templates from Concise Summaries
Horacio Saggion 270

Classification of South African languages using text and acoustic based methods: A case of six selected languages
Peleira Nicholas Zulu 280

Improving Syntax-Augmented Machine Translation by Coarsening the Label Set
Greg Hanneman and Alon Lavie 288

Keyphrase Extraction for N-best Reranking in Multi-Sentence Compression
Florian Boudin and Emmanuel Morin 298

Development of a Persian Syntactic Dependency Treebank
Mohammad Sadegh Rasooli, Manouchehr Kouhestani and Amirsaeid Moloodi 306

Improving reordering performance using higher order and structural features
Mitesh M. Khapra, Ananthakrishnan Ramanathan and Karthik Visweswariah 315

Massively Parallel Suffix Array Queries and On-Demand Phrase Extraction for Statistical Machine Translation Using GPUs
Hua He, Jimmy Lin and Adam Lopez 325

Discriminative Training of 150 Million Translation Parameters and Its Application to Pruning
Hendra Setiawan and Bowen Zhou 335

Applying Pairwise Ranked Optimisation to Improve the Interpolation of Translation Models
Barry Haddow 342

Dialectal Arabic to English Machine Translation: Pivoting through Modern Standard Arabic
Wael Salloum and Nizar Habash 348

What to do about bad language on the internet
Jacob Eisenstein 359

Minibatch and Parallelization for Online Large Margin Structured Learning
Kai Zhao and Liang Huang 370

Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters
Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider and Noah A. Smith 380

Parser lexicalisation through self-learning
Marek Rei and Ted Briscoe 391

Mining User Relations from Online Discussions using Sentiment Analysis and Probabilistic Matrix Factorization
Minghui Qiu, Liu Yang and Jing Jiang 401

Focused training sets to reduce noise in NER feature models
Amber McKenzie 411

Learning to Relate Literal and Sentimental Descriptions of Visual Properties
Mark Yatskar, Svitlana Volkova, Asli Celikyilmaz, Bill Dolan and Luke Zettlemoyer 416

Morphological Analysis and Disambiguation for Dialectal Arabic
Nizar Habash, Ryan Roth, Owen Rambow, Ramy Eskander and Nadi Tomeh 426

Using a Supertagged Dependency Language Model to Select a Good Translation in System Combination
Wei-Yun Ma and Kathleen McKeown 433

Dudley North visits North London: Learning When to Transliterate to Arabic
Mahmoud Azab, Houda Bouamor, Behrang Mohit and Kemal Oflazer 439

Better Twitter Summaries?
Joel Judd and Jugal Kalita 445

Training MRF-Based Phrase Translation Models using Gradient Ascent
Jianfeng Gao and Xiaodong He 450

Automatic Morphological Enrichment of a Morphologically Underspecified Treebank
Sarah Alkuhlani, Nizar Habash and Ryan Roth 460

A Beam-Search Decoder for Normalization of Social Media Text with Application to Machine Translation
Pidong Wang and Hwee Tou Ng 471

Parameter Estimation for LDA-Frames
Jiří Materna 482

Approximate PCFG Parsing Using Tensor Decomposition
Shay B. Cohen, Giorgio Satta and Michael Collins 487

Negative Deceptive Opinion Spam
Myle Ott, Claire Cardie and Jeffrey T. Hancock 497

Improving speech synthesis quality by reducing pitch peaks in the source recordings
Luisina Violante, Pablo Rodríguez Zivic and Agustín Gravano 502

Robust Systems for Preposition Error Correction Using Wikipedia Revisions
Aoife Cahill, Nitin Madnani, Joel Tetreault and Diane Napolitano 507

Supervised Bilingual Lexicon Induction with Multiple Monolingual Signals
Ann Irvine and Chris Callison-Burch 518

Creating Reverse Bilingual Dictionaries
Khang Nhut Lam and Jugal Kalita 524

Identification of Temporal Event Relationships in Biographical Accounts
Lucian Silcox and Emmett Tomai 529

Predicative Adjectives: An Unsupervised Criterion to Extract Subjective Adjectives
Michael Wiegand, Josef Ruppenhofer and Dietrich Klakow 534

Modeling Syntactic and Semantic Structures in Hierarchical Phrase-based Translation
Junhui Li, Philip Resnik and Hal Daumé III 540

Using Derivation Trees for Informative Treebank Inter-Annotator Agreement Evaluation
Seth Kulick, Ann Bies, Justin Mott, Mohamed Maamouri, Beatrice Santorini and Anthony Kroch 550

Embracing Ambiguity: A Comparison of Annotation Methodologies for Crowdsourcing Word Sense Labels
David Jurgens 556

Compound Embedding Features for Semi-supervised Learning
Mo Yu, Tiejun Zhao, Daxiang Dong, Hao Tian and Dianhai Yu 563

On Quality Ratings for Spoken Dialogue Systems – Experts vs. Users
Stefan Ultes, Alexander Schmitt and Wolfgang Minker 569

Overcoming the Memory Bottleneck in Distributed Training of Latent Variable Models of Text
Yi Yang, Alexander Yates and Doug Downey 579

Processing Spontaneous Orthography
Ramy Eskander, Nizar Habash, Owen Rambow and Nadi Tomeh 585

Purpose and Polarity of Citation: Towards NLP-based Bibliometrics
Amjad Abu-Jbara, Jefferson Ezra and Dragomir Radev 596

Estimating effect size across datasets
Anders Søgaard 607

Systematic Comparison of Professional and Crowdsourced Reference Translations for Machine Translation
Rabih Zbib, Gretchen Markiewicz, Spyros Matsoukas, Richard Schwartz and John Makhoul . 612

Down-stream effects of tree-to-dependency conversions
Jakob Elming, Anders Johannsen, Sigrid Klerke, Emanuele Lapponi, Hector Martinez Alonso and Anders Søgaard 617

The Life and Death of Discourse Entities: Identifying Singleton Mentions
Marta Recasens, Marie-Catherine de Marneffe and Christopher Potts 627

Automatic Generation of English Respellings
Bradley Hauer and Grzegorz Kondrak 634

A Simple, Fast, and Effective Reparameterization of IBM Model 2
Chris Dyer, Victor Chahuneau and Noah A. Smith 644

Phrase Training Based Adaptation for Statistical Machine Translation
Saab Mansour and Hermann Ney 649

Translation Acquisition Using Synonym Sets
Daniel Andrade, Masaki Tsuchida, Takashi Onishi and Kai Ishikawa 655

Supersense Tagging for Arabic: the MT-in-the-Middle Attack
Nathan Schneider, Behrang Mohit, Chris Dyer, Kemal Oflazer and Noah A. Smith 661

Zipfian corruptions for robust POS tagging
Anders Søgaard 668

A Multi-Dimensional Bayesian Approach to Lexical Style
Julian Brooke and Graeme Hirst 673

Unsupervised Domain Tuning to Improve Word Sense Disambiguation
Judita Preiss and Mark Stevenson 680

What's in a Domain? Multi-Domain Learning for Multi-Attribute Data
Mahesh Joshi, Mark Dredze, William W. Cohen and Carolyn P. Rosé 685

An opinion about opinions about opinions: subjectivity and the aggregate reader
Asad Sayeed 691

An Examination of Regret in Bullying Tweets
Jun-Ming Xu, Benjamin Burchfiel, Xiaojin Zhu and Amy Bellmore 697

A Cross-language Study on Automatic Speech Disfluency Detection
Wen Wang, Andreas Stolcke, Jiahong Yuan and Mark Liberman 703

Distributional semantic models for the evaluation of disordered language
Masoud Rouhizadeh, Emily Prud'hommeaux, Brian Roark and Jan van Santen 709

Atypical Prosodic Structure as an Indicator of Reading Level and Text Difficulty
Julie Medero and Mari Ostendorf 715

Using Document Summarization Techniques for Speech Data Subset Selection
Kai Wei, Yuzong Liu, Katrin Kirchhoff and Jeff Bilmes 721

Semi-Supervised Discriminative Language Modeling with Out-of-Domain Text Data
Arda Çelebi and Murat Saraçlar 727

More than meets the eye: Study of Human Cognition in Sense Annotation
Salil Joshi, Diptesh Kanojia and Pushpak Bhattacharyya 733

Improving Lexical Semantics for Sentential Semantics: Modeling Selectional Preference and Similar Words in a Latent Variable Model
Weiwei Guo and Mona Diab 739

Linguistic Regularities in Continuous Space Word Representations
Tomas Mikolov, Wen-tau Yih and Geoffrey Zweig 746

TruthTeller: Annotating Predicate Truth
Amnon Lotan, Asher Stern and Ido Dagan 752

PPDB: The Paraphrase Database
Juri Ganitkevitch, Benjamin Van Durme and Chris Callison-Burch 758

Exploiting the Scope of Negations and Heterogeneous Features for Relation Extraction: A Case Study for Drug-Drug Interaction Extraction
Md. Faisal Mahbub Chowdhury and Alberto Lavelli 765

Graph-Based Seed Set Expansion for Relation Extraction Using Random Walk Hitting Times
Joel Lang and James Henderson 772

Distant Supervision for Relation Extraction with an Incomplete Knowledge Base
Bonan Min, Ralph Grishman, Li Wan, Chang Wang and David Gondek 777

Measuring the Structural Importance through Rhetorical Structure Index
Narine Kokhlikyan, Alex Waibel, Yuqi Zhang and Joy Ying Zhang 783

Separating Fact from Fear: Tracking Flu Infections on Twitter
Alex Lamb, Michael J. Paul and Mark Dredze 789

Differences in User Responses to a Wizard-of-Oz versus Automated System
Jesse Thomason and Diane Litman 796

Improving the Quality of Minority Class Identification in Dialog Act Tagging
Adinoyi Omuya, Vinodkumar Prabhakaran and Owen Rambow 802

Discourse Connectors for Latent Subjectivity in Sentiment Analysis
Rakshit Trivedi and Jacob Eisenstein 808

Coherence Modeling for the Automated Assessment of Spontaneous Spoken Responses
Xinhao Wang, Keelan Evanini and Klaus Zechner 814

Disfluency Detection Using Multi-step Stacked Learning
Xian Qian and Yang Liu 820

Using Semantic Unification to Generate Regular Expressions from Natural Language
Nate Kushman and Regina Barzilay 826

Probabilistic Frame Induction
Jackie Chi Kit Cheung, Hoifung Poon and Lucy Vanderwende 837

A Quantum-Theoretic Approach to Distributional Semantics
William Blacoe, Elham Kashefi and Mirella Lapata 847

Answer Extraction as Sequence Tagging with Tree Edit Distance
Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch and Peter Clark 858

Open Information Extraction with Tree Kernels
Ying Xu, Mi-Young Kim, Kevin Quinn, Randy Goebel and Denilson Barbosa 868

Finding What Matters in Questions
Xiaoqiang Luo, Hema Raghavan, Vittorio Castelli, Sameer Maskey and Radu Florian 878

A Just-In-Time Keyword Extraction from Meeting Transcripts
Hyun-Je Song, Junho Go, Seong-Bae Park and Se-Young Park 888

Same Referent, Different Words: Unsupervised Mining of Opaque Coreferent Mentions
Marta Recasens, Matthew Can and Daniel Jurafsky 897

Global Inference for Bridging Anaphora Resolution
Yufang Hou, Katja Markert and Michael Strube 907

Classifying Temporal Relations with Rich Linguistic Knowledge
Jennifer D'Souza and Vincent Ng 918

Improved Information Structure Analysis of Scientific Documents Through Discourse and Lexical Constraints
Yufan Guo, Roi Reichart and Anna Korhonen 928

Adaptation of Reordering Models for Statistical Machine Translation
Boxing Chen, George Foster and Roland Kuhn 938

Multi-Metric Optimization Using Ensemble Tuning
Baskaran Sankaran, Anoop Sarkar and Kevin Duh 947

Grouping Language Model Boundary Words to Speed K–Best Extraction from Hypergraphs
Kenneth Heafield, Philipp Koehn and Alon Lavie 958

A Systematic Bayesian Treatment of the IBM Alignment Models
Yarin Gal and Phil Blunsom 969

Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering
Ekaterina Shutova and Lin Sun 978

Three Knowledge-Free Methods for Automatic Lexical Chain Extraction
Steffen Remus and Chris Biemann 989

Combining Heterogeneous Models for Measuring Relational Similarity
Alisa Zhila, Wen-tau Yih, Christopher Meek, Geoffrey Zweig and Tomas Mikolov 1000

Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter
Shane Bergsma, Mark Dredze, Benjamin Van Durme, Theresa Wilson and David Yarowsky 1010

To Link or Not to Link? A Study on End-to-End Tweet Entity Linking
Stephen Guo, Ming-Wei Chang and Emre Kiciman 1020

A Latent Variable Model for Viewpoint Discovery from Threaded Forum Posts
Minghui Qiu and Jing Jiang 1031

Identifying Intention Posts in Discussion Forums
Zhiyuan Chen, Bing Liu, Meichun Hsu, Malu Castellanos and Riddhiman Ghosh 1041

Dependency-based empty category detection via phrase structure trees
Nianwen Xue and Yaqin Yang 1051

Target Language Adaptation of Discriminative Transfer Parsers
Oscar Täckström, Ryan McDonald and Joakim Nivre 1061

Emergence of Gricean Maxims from Multi-Agent Decision Theory
Adam Vogel, Max Bodoia, Christopher Potts and Daniel Jurafsky 1072

Open Dialogue Management for Relational Databases
Ben Hixon and Rebecca J. Passonneau 1082

A method for the approximation of incremental understanding of explicit utterance meaning using predictive models in finite domains
David DeVault and David Traum 1092

Paving the Way to a Large-scale Pseudosense-annotated Dataset
Mohammad Taher Pilehvar and Roberto Navigli 1100

Labeling the Languages of Words in Mixed-Language Documents using Weakly Supervised Methods
Ben King and Steven Abney 1110

Learning Whom to Trust with MACE
Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani and Eduard Hovy 1120

Supervised All-Words Lexical Substitution using Delexicalized Features
György Szarvas, Chris Biemann and Iryna Gurevych 1131

A Tensor-based Factorization Model of Semantic Compositionality
Tim Van de Cruys, Thierry Poibeau and Anna Korhonen 1142

A Participant-based Approach for Event Summarization Using Twitter Streams
Chao Shen, Fei Liu, Fuliang Weng and Tao Li 1152

Towards Coherent Multi-Document Summarization
Janara Christensen, Mausam, Stephen Soderland and Oren Etzioni 1163

Generating Expressions that Refer to Visible Objects
Margaret Mitchell, Kees van Deemter and Ehud Reiter 1174

Supervised Learning of Complete Morphological Paradigms
Greg Durrett and John DeNero 1185

Optimal Data Set Selection: An Application to Grapheme-to-Phoneme Conversion
Young-Bum Kim and Benjamin Snyder 1196

Knowledge-Rich Morphological Priors for Bayesian Language Models
Victor Chahuneau, Noah A. Smith and Chris Dyer 1206

Conference Program

Monday, June 10, 2013

8:45–9:00 Welcome to NAACL 2013!

9:00–10:10 Invited talk by Gina Kuperberg – Predicting Meaning: What the Brain tells us about the Architecture of Language Comprehension

10:10–10:40 Break

M1a: Machine Translation

10:40-11:05 *Model With Minimal Translation Units, But Decode With Phrases*
Nadir Durrani, Alexander Fraser and Helmut Schmid

11:05-11:30 *Beyond Left-to-Right: Multiple Decomposition Structures for SMT*
Hui Zhang, Kristina Toutanova, Chris Quirk and Jianfeng Gao

11:30-11:55 *Improved Reordering for Phrase-Based Translation using Sparse Features*
Colin Cherry

11:55-12:20 *Simultaneous Word-Morpheme Alignment for Statistical Machine Translation*
Elif Eyigöz, Daniel Gildea and Kemal Oflazer

M1b: Information Extraction

10:40-11:05 *Multi-faceted Event Recognition with Bootstrapped Dictionaries*
Ruihong Huang and Ellen Riloff

11:05-11:30 *Named Entity Recognition with Bilingual Constraints*
Wanxiang Che, Mengqiu Wang, Christopher D. Manning and Ting Liu

11:30-11:55 *Minimally Supervised Method for Multilingual Paraphrase Extraction from Definition Sentences on the Web*
Yulan Yan, Chikara Hashimoto, Kentaro Torisawa, Takao Kawai, Jun'ichi Kazama and Stijn De Saeger

11:55-12:20 *Relation Extraction with Matrix Factorization and Universal Schemas*
Sebastian Riedel, Limin Yao, Andrew McCallum and Benjamin M. Marlin

Monday, June 10, 2013 (continued)

M1c: Cognitive and Psycholinguistics

10:40-11:05 *Extracting the Native Language Signal for Second Language Acquisition*
Ben Swanson and Eugene Charniak

11:05-11:30 *An Analysis of Frequency- and Memory-Based Processing Costs*
Marten van Schijndel and William Schuler

11:30-11:55 *Cross-Lingual Semantic Similarity of Words as the Similarity of Their Semantic Word Responses*
Ivan Vulić and Marie-Francine Moens

11:55-12:20 *Combining multiple information types in Bayesian word segmentation*
Gabriel Doyle and Roger Levy

12:20–2:00 Lunch

M2a: Parsing and Syntax

2:00-2:25 *Training Parsers on Incompatible Treebanks*
Richard Johansson

2:25-2:50 *Learning a Part-of-Speech Tagger from Two Hours of Annotation*
Dan Garrette and Jason Baldridge

2:50-3:15 *Experiments with Spectral Learning of Latent-Variable PCFGs*
Shay B. Cohen, Karl Stratos, Michael Collins, Dean P. Foster and Lyle Ungar

Monday, June 10, 2013 (continued)

M2b: Topic Modeling and Text Mining

2:00-2:25 *Representing Topics Using Images*
Nikolaos Aletras and Mark Stevenson

2:25-2:50 *Drug Extraction from the Web: Summarizing Drug Experiences with Multi-Dimensional Topic Models*
Michael J. Paul and Mark Dredze

2:50-3:15 *Towards Topic Labeling with Phrase Entailment and Aggregation*
Yashar Mehdad, Giuseppe Carenini, Raymond T. Ng and Shafiq Joty

3:15-3:40 *Topic Segmentation with a Structured Topic Model*
Lan Du, Wray Buntine and Mark Johnson

M2c: Spoken Language Processing

2:00-2:25 *Text Alignment for Real-Time Crowd Captioning*
Iftekhar Naim, Daniel Gildea, Walter Lasecki and Jeffrey P. Bigham

2:25-2:50 *Discriminative Joint Modeling of Lexical Variation and Acoustic Confusion for Automated Narrative Retelling Assessment*
Maider Lehr, Izhak Shafran, Emily Prud'hommeaux and Brian Roark

2:50-3:15 *Using Out-of-Domain Data for Lexical Addressee Detection in Human-Human-Computer Dialog*
Heeyoung Lee, Andreas Stolcke and Elizabeth Shriberg

3:15-3:40 *Segmentation Strategies for Streaming Speech Translation*
Vivek Kumar Rangarajan Sridhar, John Chen, Srinivas Bangalore, Andrej Ljolje and Rathinavelu Chengalvarayan

3:40–4:10 Break

4:10–6:00 Poster madness!

Enforcing Subcategorization Constraints in a Parser Using Sub-parses Recombining
Seyed Abolghasem Mirroshandel, Alexis Nasr and Benoît Sagot

Large-Scale Discriminative Training for Statistical Machine Translation Using Held-Out Line Search
Jeffrey Flanigan, Chris Dyer and Jaime Carbonell

Monday, June 10, 2013 (continued)

Measuring Term Informativeness in Context
Zhaohui Wu and C. Lee Giles

Unsupervised Learning Summarization Templates from Concise Summaries
Horacio Saggion

Classification of South African languages using text and acoustic based methods: A case of six selected languages
Peleira Nicholas Zulu

Improving Syntax-Augmented Machine Translation by Coarsening the Label Set
Greg Hanneman and Alon Lavie

Keyphrase Extraction for N-best Reranking in Multi-Sentence Compression
Florian Boudin and Emmanuel Morin

Development of a Persian Syntactic Dependency Treebank
Mohammad Sadegh Rasooli, Manouchehr Kouhestani and Amirsaeid Moloodi

Improving reordering performance using higher order and structural features
Mitesh M. Khapra, Ananthakrishnan Ramanathan and Karthik Visweswariah

Massively Parallel Suffix Array Queries and On-Demand Phrase Extraction for Statistical Machine Translation Using GPUs
Hua He, Jimmy Lin and Adam Lopez

Discriminative Training of 150 Million Translation Parameters and Its Application to Pruning
Hendra Setiawan and Bowen Zhou

Applying Pairwise Ranked Optimisation to Improve the Interpolation of Translation Models
Barry Haddow

Dialectal Arabic to English Machine Translation: Pivoting through Modern Standard Arabic
Wael Salloum and Nizar Habash

What to do about bad language on the internet
Jacob Eisenstein

Minibatch and Parallelization for Online Large Margin Structured Learning
Kai Zhao and Liang Huang

Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters
Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider and Noah A. Smith

Parser lexicalisation through self-learning
Marek Rei and Ted Briscoe

Mining User Relations from Online Discussions using Sentiment Analysis and Probabilistic Matrix Factorization
Minghui Qiu, Liu Yang and Jing Jiang

Focused training sets to reduce noise in NER feature models
Amber McKenzie

Learning to Relate Literal and Sentimental Descriptions of Visual Properties
Mark Yatskar, Svitlana Volkova, Asli Celikyilmaz, Bill Dolan and Luke Zettlemoyer

Morphological Analysis and Disambiguation for Dialectal Arabic
Nizar Habash, Ryan Roth, Owen Rambow, Ramy Eskander and Nadi Tomeh

Using a Supertagged Dependency Language Model to Select a Good Translation in System Combination
Wei-Yun Ma and Kathleen McKeown

Dudley North visits North London: Learning When to Transliterate to Arabic
Mahmoud Azab, Houda Bouamor, Behrang Mohit and Kemal Oflazer

Better Twitter Summaries?
Joel Judd and Jugal Kalita

Training MRF-Based Phrase Translation Models using Gradient Ascent
Jianfeng Gao and Xiaodong He

Automatic Morphological Enrichment of a Morphologically Underspecified Treebank
Sarah Alkuhlani, Nizar Habash and Ryan Roth

A Beam-Search Decoder for Normalization of Social Media Text with Application to Machine Translation
Pidong Wang and Hwee Tou Ng

Parameter Estimation for LDA-Frames
Jiří Materna

Approximate PCFG Parsing Using Tensor Decomposition
Shay B. Cohen, Giorgio Satta and Michael Collins

Negative Deceptive Opinion Spam
Myle Ott, Claire Cardie and Jeffrey T. Hancock

Improving speech synthesis quality by reducing pitch peaks in the source recordings
Luisina Violante, Pablo Rodríguez Zivic and Agustín Gravano

Robust Systems for Preposition Error Correction Using Wikipedia Revisions
Aoife Cahill, Nitin Madnani, Joel Tetreault and Diane Napolitano

Supervised Bilingual Lexicon Induction with Multiple Monolingual Signals
Ann Irvine and Chris Callison-Burch

Creating Reverse Bilingual Dictionaries
Khang Nhut Lam and Jugal Kalita

Identification of Temporal Event Relationships in Biographical Accounts
Lucian Silcox and Emmett Tomai

Predicative Adjectives: An Unsupervised Criterion to Extract Subjective Adjectives
Michael Wiegand, Josef Ruppenhofer and Dietrich Klakow

Modeling Syntactic and Semantic Structures in Hierarchical Phrase-based Translation
Junhui Li, Philip Resnik and Hal Daumé III

Using Derivation Trees for Informative Treebank Inter-Annotator Agreement Evaluation
Seth Kulick, Ann Bies, Justin Mott, Mohamed Maamouri, Beatrice Santorini and Anthony Kroch

Embracing Ambiguity: A Comparison of Annotation Methodologies for Crowdsourcing Word Sense Labels
David Jurgens

Compound Embedding Features for Semi-supervised Learning
Mo Yu, Tiejun Zhao, Daxiang Dong, Hao Tian and Dianhai Yu

On Quality Ratings for Spoken Dialogue Systems – Experts vs. Users
Stefan Ultes, Alexander Schmitt and Wolfgang Minker

Overcoming the Memory Bottleneck in Distributed Training of Latent Variable Models of Text
Yi Yang, Alexander Yates and Doug Downey

Processing Spontaneous Orthography
Ramy Eskander, Nizar Habash, Owen Rambow and Nadi Tomeh

Purpose and Polarity of Citation: Towards NLP-based Bibliometrics
Amjad Abu-Jbara, Jefferson Ezra and Dragomir Radev

Estimating effect size across datasets
Anders Søgaard

Systematic Comparison of Professional and Crowdsourced Reference Translations for Machine Translation
Rabih Zbib, Gretchen Markiewicz, Spyros Matsoukas, Richard Schwartz and John Makhoul

Down-stream effects of tree-to-dependency conversions
Jakob Elming, Anders Johannsen, Sigrid Klerke, Emanuele Lapponi, Hector Martinez Alonso and Anders Søgaard

6:00–6:30 Break

6:30–8:30 Poster and Demonstrations Session

Tuesday, June 11, 2013

9:15–9:25 Best paper awards

Best Short Paper

9:25–9:45 *The Life and Death of Discourse Entities: Identifying Singleton Mentions*
Marta Recasens, Marie-Catherine de Marneffe and Christopher Potts

IBM Best Student Paper

9:45–10:15 *Automatic Generation of English Respellings*
Bradley Hauer and Grzegorz Kondrak

10:15–10:45 Break

T1a: Machine Translation and Multilinguality

10:45-11:00 *A Simple, Fast, and Effective Reparameterization of IBM Model 2*
Chris Dyer, Victor Chahuneau and Noah A. Smith

11:00-11:15 *Phrase Training Based Adaptation for Statistical Machine Translation*
Saab Mansour and Hermann Ney

11:15-11:30 *Translation Acquisition Using Synonym Sets*
Daniel Andrade, Masaki Tsuchida, Takashi Onishi and Kai Ishikawa

11:30-11:45 *Supersense Tagging for Arabic: the MT-in-the-Middle Attack*
Nathan Schneider, Behrang Mohit, Chris Dyer, Kemal Oflazer and Noah A. Smith

11:45-12:00 *Zipfian corruptions for robust POS tagging*
Anders Søgaard

Tuesday, June 11, 2013 (continued)

T1b: Sentiment Analysis and Topic Modeling

10:45-11:00 *A Multi-Dimensional Bayesian Approach to Lexical Style*
Julian Brooke and Graeme Hirst

11:00-11:15 *Unsupervised Domain Tuning to Improve Word Sense Disambiguation*
Judita Preiss and Mark Stevenson

11:15-11:30 *What's in a Domain? Multi-Domain Learning for Multi-Attribute Data*
Mahesh Joshi, Mark Dredze, William W. Cohen and Carolyn P. Rosé

11:30-11:45 *An opinion about opinions about opinions: subjectivity and the aggregate reader*
Asad Sayeed

11:45-12:00 *An Examination of Regret in Bullying Tweets*
Jun-Ming Xu, Benjamin Burchfiel, Xiaojin Zhu and Amy Bellmore

T1c: Spoken Language Processing

10:45-11:00 *A Cross-language Study on Automatic Speech Disfluency Detection*
Wen Wang, Andreas Stolcke, Jiahong Yuan and Mark Liberman

11:00-11:15 *Distributional semantic models for the evaluation of disordered language*
Masoud Rouhizadeh, Emily Prud'hommeaux, Brian Roark and Jan van Santen

11:15-11:30 *Atypical Prosodic Structure as an Indicator of Reading Level and Text Difficulty*
Julie Medero and Mari Ostendorf

11:30-11:45 *Using Document Summarization Techniques for Speech Data Subset Selection*
Kai Wei, Yuzong Liu, Katrin Kirchhoff and Jeff Bilmes

11:45-12:00 *Semi-Supervised Discriminative Language Modeling with Out-of-Domain Text Data*
Arda Çelebi and Murat Saraçlar

12:00–2:00 Lunch (and Business Meeting 1-2)

Tuesday, June 11, 2013 (continued)

T2a: Semantics

2:00-2:15 *More than meets the eye: Study of Human Cognition in Sense Annotation*
Salil Joshi, Diptesh Kanojia and Pushpak Bhattacharyya

2:15-2:30 *Improving Lexical Semantics for Sentential Semantics: Modeling Selectional Preference and Similar Words in a Latent Variable Model*
Weiwei Guo and Mona Diab

2:30-2:45 *Linguistic Regularities in Continuous Space Word Representations*
Tomas Mikolov, Wen-tau Yih and Geoffrey Zweig

2:45-3:00 *TruthTeller: Annotating Predicate Truth*
Amnon Lotan, Asher Stern and Ido Dagan

3:00-3:15 *PPDB: The Paraphrase Database*
Juri Ganitkevitch, Benjamin Van Durme and Chris Callison-Burch

T2b: Information Extraction

2:00-2:15 *Exploiting the Scope of Negations and Heterogeneous Features for Relation Extraction: A Case Study for Drug-Drug Interaction Extraction*
Md. Faisal Mahbub Chowdhury and Alberto Lavelli

2:15-2:30 *Graph-Based Seed Set Expansion for Relation Extraction Using Random Walk Hitting Times*
Joel Lang and James Henderson

2:30-2:45 *Distant Supervision for Relation Extraction with an Incomplete Knowledge Base*
Bonan Min, Ralph Grishman, Li Wan, Chang Wang and David Gondek

2:45-3:00 *Measuring the Structural Importance through Rhetorical Structure Index*
Narine Kokhlikyan, Alex Waibel, Yuqi Zhang and Joy Ying Zhang

3:00-3:15 *Separating Fact from Fear: Tracking Flu Infections on Twitter*
Alex Lamb, Michael J. Paul and Mark Dredze

Tuesday, June 11, 2013 (continued)

T2c: Discourse and Dialog

2:00-2:15 *Differences in User Responses to a Wizard-of-Oz versus Automated System*
Jesse Thomason and Diane Litman

2:15-2:30 *Improving the Quality of Minority Class Identification in Dialog Act Tagging*
Adinoyi Omuya, Vinodkumar Prabhakaran and Owen Rambow

2:30-2:45 *Discourse Connectors for Latent Subjectivity in Sentiment Analysis*
Rakshit Trivedi and Jacob Eisenstein

2:45-3:00 *Coherence Modeling for the Automated Assessment of Spontaneous Spoken Responses*
Xinhao Wang, Keelan Evanini and Klaus Zechner

3:00-3:15 *Disfluency Detection Using Multi-step Stacked Learning*
Xian Qian and Yang Liu

3:15–3:45 Break

T3a: Semantics

4:10-4:35 *Using Semantic Unification to Generate Regular Expressions from Natural Language*
Nate Kushman and Regina Barzilay

4:35-5:00 *Probabilistic Frame Induction*
Jackie Chi Kit Cheung, Hoifung Poon and Lucy Vanderwende

5:00–5:25 *A Quantum-Theoretic Approach to Distributional Semantics*
William Blacoe, Elham Kashefi and Mirella Lapata

Tuesday, June 11, 2013 (continued)

T3b: Information Extraction

3:45-4:10 *Answer Extraction as Sequence Tagging with Tree Edit Distance*
Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch and Peter Clark

4:10-4:35 *Open Information Extraction with Tree Kernels*
Ying Xu, Mi-Young Kim, Kevin Quinn, Randy Goebel and Denilson Barbosa

4:35-5:00 *Finding What Matters in Questions*
Xiaoqiang Luo, Hema Raghavan, Vittorio Castelli, Sameer Maskey and Radu Florian

5:00-5:25 *A Just-In-Time Keyword Extraction from Meeting Transcripts*
Hyun-Je Song, Junho Go, Seong-Bae Park and Se-Young Park

T3c: Discourse

3:45-4:10 *Same Referent, Different Words: Unsupervised Mining of Opaque Coreferent Mentions*
Marta Recasens, Matthew Can and Daniel Jurafsky

4:10-4:35 *Global Inference for Bridging Anaphora Resolution*
Yufang Hou, Katja Markert and Michael Strube

4:35-5:00 *Classifying Temporal Relations with Rich Linguistic Knowledge*
Jennifer D'Souza and Vincent Ng

5:00-5:25 *Improved Information Structure Analysis of Scientific Documents Through Discourse and Lexical Constraints*
Yufan Guo, Roi Reichart and Anna Korhonen

7:00–9:30 Banquet

Wednesday, June 12, 2013

9:00–10:10 Invited talk by Kathleen McKeown – Natural Language Applications from Fact to Fiction

10:10–10:40 Break

W1a: Machine Translation

10:40-11:05 *Adaptation of Reordering Models for Statistical Machine Translation*
Boxing Chen, George Foster and Roland Kuhn

11:05-11:30 *Multi-Metric Optimization Using Ensemble Tuning*
Baskaran Sankaran, Anoop Sarkar and Kevin Duh

11:30-11:55 *Grouping Language Model Boundary Words to Speed K–Best Extraction from Hypergraphs*
Kenneth Heafield, Philipp Koehn and Alon Lavie

11:55-12:20 *A Systematic Bayesian Treatment of the IBM Alignment Models*
Yarin Gal and Phil Blunsom

W1b: Semantics

10:40-11:05 *Unsupervised Metaphor Identification Using Hierarchical Graph Factorization Clustering*
Ekaterina Shutova and Lin Sun

11:05-11:30 *Three Knowledge-Free Methods for Automatic Lexical Chain Extraction*
Steffen Remus and Chris Biemann

11:30-11:55 *Combining Heterogeneous Models for Measuring Relational Similarity*
Alisa Zhila, Wen-tau Yih, Christopher Meek, Geoffrey Zweig and Tomas Mikolov

Wednesday, June 12, 2013 (continued)

W1c: Social Media Processing

10:40-11:05 *Broadly Improving User Classification via Communication-Based Name and Location Clustering on Twitter*
Shane Bergsma, Mark Dredze, Benjamin Van Durme, Theresa Wilson and David Yarowsky

11:05-11:30 *To Link or Not to Link? A Study on End-to-End Tweet Entity Linking*
Stephen Guo, Ming-Wei Chang and Emre Kiciman

11:30-11:55 *A Latent Variable Model for Viewpoint Discovery from Threaded Forum Posts*
Minghui Qiu and Jing Jiang

11:55-12:20 *Identifying Intention Posts in Discussion Forums*
Zhiyuan Chen, Bing Liu, Meichun Hsu, Malu Castellanos and Riddhiman Ghosh

12:20–2:00 Lunch

W2a: Parsing and Syntax

2:25-2:50 *Dependency-based empty category detection via phrase structure trees*
Nianwen Xue and Yaqin Yang

2:50-3:15 *Target Language Adaptation of Discriminative Transfer Parsers*
Oscar Täckström, Ryan McDonald and Joakim Nivre

W2b: Dialog

2:00-2:25 *Emergence of Gricean Maxims from Multi-Agent Decision Theory*
Adam Vogel, Max Bodoia, Christopher Potts and Daniel Jurafsky

2:25-2:50 *Open Dialogue Management for Relational Databases*
Ben Hixon and Rebecca J. Passonneau

2:50-3:15 *A method for the approximation of incremental understanding of explicit utterance meaning using predictive models in finite domains*
David DeVault and David Traum

Wednesday, June 12, 2013 (continued)

W2c: Annotation and Language Resources

2:00-2:25 *Paving the Way to a Large-scale Pseudosense-annotated Dataset*
Mohammad Taher Pilehvar and Roberto Navigli

2:25-2:50 *Labeling the Languages of Words in Mixed-Language Documents using Weakly Supervised Methods*
Ben King and Steven Abney

2:50-3:15 *Learning Whom to Trust with MACE*
Dirk Hovy, Taylor Berg-Kirkpatrick, Ashish Vaswani and Eduard Hovy

3:15–3:45 Break

W3a: Semantics and Syntax

3:45-4:10 *Supervised All-Words Lexical Substitution using Delexicalized Features*
György Szarvas, Chris Biemann and Iryna Gurevych

4:10-4:35 *A Tensor-based Factorization Model of Semantic Compositionality*
Tim Van de Cruys, Thierry Poibeau and Anna Korhonen

W3b: Summarization and Generation

3:45-4:10 *A Participant-based Approach for Event Summarization Using Twitter Streams*
Chao Shen, Fei Liu, Fuliang Weng and Tao Li

4:10-4:35 *Towards Coherent Multi-Document Summarization*
Janara Christensen, Mausam, Stephen Soderland and Oren Etzioni

4:35-5:00 *Generating Expressions that Refer to Visible Objects*
Margaret Mitchell, Kees van Deemter and Ehud Reiter

Wednesday, June 12, 2013 (continued)

W3c: Morphology and Phonology

3:45-4:10 *Supervised Learning of Complete Morphological Paradigms*
Greg Durrett and John DeNero

4:10-4:35 *Optimal Data Set Selection: An Application to Grapheme-to-Phoneme Conversion*
Young-Bum Kim and Benjamin Snyder

4:35-5:00 *Knowledge-Rich Morphological Priors for Bayesian Language Models*
Victor Chahuneau, Noah A. Smith and Chris Dyer