NAACL HLT 2012

The 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

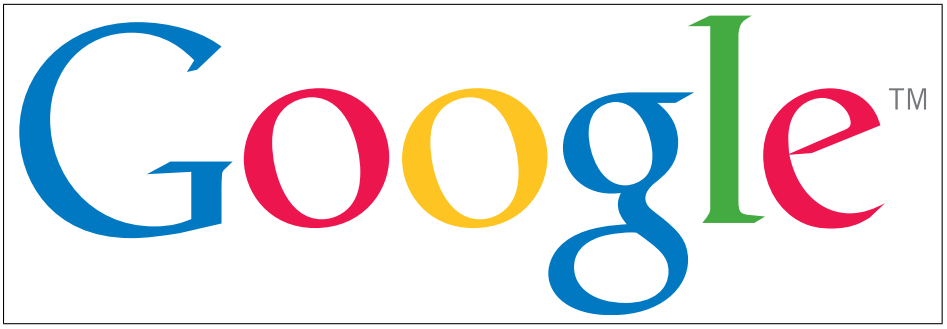
Proceedings of the Conference

June 3-8, 2012
Montréal, Canada

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53707
USA

PLATINUM SPONSORS:

GOLD SPONSORS:



SILVER SPONSORS:

BRONZE SPONSORS:

SUPPORTERS:

©2012 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

ISBN 978-1-937284-20-6 / 1-937284-20-4

Preface: General Chair

It is my pleasure and honor to welcome you to the 2012 NAACL Human Language Technologies Conference in beautiful Montreal, Canada (in our Canadian spirit, let me add, bienvenue!). The organizing committee has put in a great deal of effort on the programs in the upcoming week. I hope you will enjoy what the conference has to offer!

The core of the conference is the 3-day main technical program consisting of oral and poster presentations of papers, keynote addresses, and a novel "NLP Idol" session. I am very fortunate to have this key piece of the conference in the hands of Eric Fosler-Lussier, Ellen Riloff, and Srinivas Bangalore, three extremely dedicated and capable program co-chairs. Eric, Ellen, and Srini managed the whole process so well that I realized early on in the conference-planning process that I could leave this important part of the conference entirely in their hands. The main technical program that you will see here is the fruit of their labor, with the assistance of 22 area chairs and hundreds of people who reviewed or submitted papers, or both. Thank you all!

To be held in conjunction with the main conference poster session is the Demonstrations program. I would like to thank Aria Haghighi and Yaser Al-Onaizan for selecting a dozen interesting system demos as part of this program. Another component of the poster session is presentations from the Student Research Workshop. I am grateful to co-chairs Rivka Levitan and Myle Ott as well as faculty advisors Roger Levy and Ani Nenkova for identifying our rising stars, and for organizing a roundtable discussion on peer review standards and practices for all participants.

For the Tutorials program, I thank Jacob Eisenstein and Radu Florian for managing the submission and reviewing process to identify 8 diverse and interesting half-day tutorials for presentation. Continuing a new tradition, the Workshops program is coordinated among the EACL, NAACL, and ACL conferences, with a joint submission and review process. I am indebted to the NAACL Workshop co-chairs Colin Cherry and Mona Diab who worked as part of the *ACL workshop committee to select a strong suite of 16 workshops for NAACL and put in significant effort working with the workshop organizers to bring the whole program together. Thank you also to the organizers of the *SEM conference, Eneko Agirre, Johan Bos, and Mona Diab, for choosing to collocate their first conference with NAACL HLT.

The USB key that contains the entire proceedings of this conference is the production of Publications co-chairs Nizar Habash and William Schuler. Special thanks go to them for their efforts in assembling all the materials and working to keep everyone on schedule for the production of the proceedings.

I would like to thank Publicity chair Smaranda Muresan for her efforts in helping to attract submissions and attendees from various communities, and Exhibits chair Joel Tetreault for helping to arrange show-and-tell space for our sponsors and other exhibitors. The website that you undoubtedly consulted countless times before getting to Montreal has been designed and maintained by Dirk Hovy. A heartfelt thank you to Dirk for taking on this task, and to Lucy Roark for designing the NAACL HLT 2012 logo.

For the financial aspects of the conference, I would like to thank our corporate, academic, and government sponsors for their contributions, and our North American sponsorship co-chairs Michael Gamon and Patrick Pantel for the role they played.

The NAACL executive board has been very supportive and helpful during the planning of this

conference. I am very grateful for the guidance and suggestions the board members, especially the two chairs Rebecca Hwa and Chris Callison-Burch, have provided.

Finally, the conference would not be happening without the expertise and dedication of Priscilla Rasmussen, ACL's business manager and this conference's Local Arrangements Chair. Working with local advisory committee members Sabine Bergler and Guy Lapalme, Priscilla is taking on local arrangements responsibilities usually split among multiple faculty members and doing a fabulous job. With her vast array of experience in every aspect of organizing a *ACL conference, Priscilla has been my go-to person and life-saver for the last 9 months. I just want to say, thank you Priscilla, it's been great working with you!

For those of you who made it this far, here's a conference-appropriate pun for your amusement.

Q: What do linguists call Santa's elves?
A: Subordinated Clauses.

Enjoy the conference!

Jennifer Chu-Carroll
IBM T.J. Watson Research Center
NAACL HLT 2012 General Chair

Preface: Program Chairs

Welcome to NAACL HLT 2012! This year's conference brings to Montreal an exciting array of work ranging across the human language technology disciplines. The main conference features both oral and poster sessions for full and short papers; the main conference is preceded by eight tutorials and followed by sixteen separate workshops as well as the First Joint Conference on Lexical and Computational Semantics (*SEM).

Within the main program, we are pleased to announce several special events. We have two excellent invited speakers starting off two days of our program. On Monday morning, we will hear from Eduard Hovy, Director of the Human Language Technology Group, Information Sciences Institute of the University of Southern California, who will speak about "A New Semantics: Merging Propositional and Distributional Information." Wednesday morning's invited speaker is James W. Pennebaker, Centennial Liberal Arts Professor and Chair of Psychology at the University of Texas at Austin; his talk is entitled "A, is, I, and, the: How our smallest words reveal the most about who we are."

Two special discussion-oriented events are also planned — on Monday during the lunch hour, the Student Research Workshop will be hosting a panel discussion on "Reviewing Practices," which is open to all conference participants. We will also have a special fun event Tuesday afternoon called "NLP Idol," where senior researchers will try to convince a panel of judges (and the audience!) that we should be paying attention to a forgotten line of research from the past by presenting papers "plucked from obscurity."

This year, 196 full papers were submitted to the conference, with 61 papers being accepted (a 31% acceptance rate); 105 short papers were submitted, with 36 acceptances (34% acceptance). The breakdown of papers by area of submission (based on author designation) and acceptances (in parentheses) were as follows:

Author-assigned Paper Category	**# Full Papers**	**# Short Papers**
Discourse, Dialogue, and Pragmatics	14 (7)	8 (3)
Document Categorization / Topic Clustering	13 (1)	7 (2)
End-to-end Language Processing Systems	4 (3)	4 (3)
Information Extraction	17 (4)	6 (0)
Information Retrieval and Question Answering	7 (3)	5 (0)
Language Resources, Novel Evaluation Methods	8 (5)	8 (3)
Machine Learning for Language Processing	21 (11)	8 (2)
Machine Translation	26 (8)	18 (5)
Phonology and Morphology, Word Segmentation	7 (1)	5 (3)
Semantics	25 (5)	7 (3)
Sentiment Analysis and Opinion Mining	12 (1)	8 (3)
Social Media Analysis and Processing	7 (2)	3 (1)
Spoken Language Processing	8 (2)	4 (2)
Summarization and Generation	8 (1)	5 (3)
Syntactic Tagging and Chunking	3 (1)	1 (0)
Syntax and Parsing	16 (6)	8 (3)

(As part of the review and assignment process, some of the papers were recategorized by the program chairs, so the acceptance numbers based on author categorization do not necessarily match the assignment of papers in the program.)

The oral and poster slots were allocated based on the suggestions of reviewers and area chairs for appropriate presentation style; both presentation types carry the same status. Fourteen full and seven short papers will be presented during an evening poster session, with buffet dinner, in conjunction with the Demo session and the Student Research Workshop posters. Preceding the poster session will be a reprise of the one-minute madness session introduced at NAACL HLT 2010, in which attendees can see an overview of the poster presentations.

Oral sessions will be held in three parallel sessions on Monday and Wednesday, with four parallel sessions on Tuesday. We have expanded presentation times to 30-minute slots for full papers, and 20-minute slots for short papers, to facilitate more discussion of 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 fine work.

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 476 reviewers 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. Constructing the program would not have been possible without 22 excellent area chairs forming the Senior Program Committee: Roberto Basili, Guiseppe Carenini, Yejin Choi, Christine Doran, Jason Eisner, George Foster, Roxana Girju, Heng Ji, Sadao Kurohashi, Matt Lease, Diane Litman, Deepak Ravichandran, Giuseppe Riccardi, Richard Rose, Giorgio Satta, Fei Sha, Suzanne Stevenson, David Traum, Scott Yih, Luke Zettlemoyer, Bowen Zhou, and Jerry Zhu. Area chairs were responsible for recruiting reviewers, 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.

The Best Paper Award session starts off Tuesday morning; this year we are pleased to present three awards — for best full paper, best short paper, and best student paper. This year's winners are:

- Best Full Paper Award: **Vine Pruning for Efficient Multi-Pass Dependency Parsing,** Alexander Rush and Slav Petrov
- Best Short Paper Award: **Trait-Based Hypothesis Selection For Machine Translation,** Jacob Devlin and Spyros Matsoukas
- IBM Best Student Paper Award: **Cross-lingual Word Clusters for Direct Transfer of Linguistic Structure,** Oscar Täckström, Ryan McDonald, Jakob Uszkoreit

We would like to thank reviewers and area chairs for nominating the best paper candidates. A subset of area chairs with expertise in the areas of the nominated papers were invaluable in helping the program chairs in the decision process. In particular, we would like to thank Christine Doran, Jason Eisner, George Foster, Diane Litman, Giorgio Satta, Luke Zettlemoyer, and Bowen Zhou for their assistance

in the decision process. We would like to note that the Best Full Paper and IBM Best Student Paper awardees both have students as first authors. The authors will be presented with a certificate and cash prize at the opening of Tuesday's session. We gratefully acknowledge IBM's support for the Student Best Paper Award.

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. Nizar Habash and William Schuler, 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 Rivka Levitan, Myle Ott, Roger Levy, and Ani Nenkova (Student Research Workshop); Jacob Eisenstein and Radu Florian (Tutorial Chairs); Colin Cherry and Mona Diab (Workshop Chairs); and Aria Haghighi and Yaser Al-Onaizan (Demo Chairs). We would also like to thank Chris Callison-Burch, Rebecca Hwa, and the NAACL Executive Board for guidance during the process. Dirk Hovy was also a valuable team member in helping us disseminate information as Webmaster.

Deserving special mention is the ever-unflappable Priscilla Rasmussen, who is doing double duty this conference as local arrangements chair and general business manager. Priscilla makes everything she is involved with go more smoothly, and we have relied on her advice greatly during the run-up to the conference.

Finally, we would like to thank our General Chair, Jennifer Chu-Carroll, for entrusting us with this job, for walking us through some of the more sticky moments, and for being a great sounding board for different ideas. In particular, her guidance was crucial in developing the concept for the NLP Idol session.

We hope that you enjoy the conference!

Eric Fosler-Lussier, The Ohio State University
Ellen Riloff, University of Utah
Srinivas Bangalore, AT&T Research

General Chair:

Jennifer Chu-Carroll, IBM T.J. Watson Research Center

Program Co-Chairs:

Eric Fosler-Lussier, The Ohio State University
Ellen Riloff, University of Utah
Srinivas Bangalore, AT&T

Area Chairs:

Roberto Basili, University of Rome: Semantics
Guiseppe Carenini, University of British Columbia: Summarization and Generation
Yejin Choi, Stony Brook University: Sentiment Analysis and Opinion Mining
Christy Doran, MITRE: Language Resources Novel Evaluation Methods
Jason Eisner, Johns Hopkins University: Phonology and Morphology Word Segmentation
George Foster, National Research Council Canada: Machine Translation
Roxana Girju University of Illinois Urbana-Champaign: Social Media Analysis and Processing
Heng Ji, City University of New York: Information Extraction
Sadao Kurohashi, Kyoto University: Syntactic Tagging and Chunking
Matt Lease, University of Texas Austin: Information Retrieval and Question Answering
Diane Litman, University of Pittsburgh: Discourse Dialogue and Pragmatics
Deepak Ravichandran, Google: Information Extraction
Guiseppe Ricardi, University of Trento: Spoken Language Processing
Richard Rose, McGill University: Spoken Language Processing
Giorgio Satta, University of Padova: Syntax and Parsing
Fei Sha, University of Southern California: Machine Learning for Language Processing
Suzanne Stevenson, University of Toronto: Semantics
David Traum, University of Southern California: End-to-end Language Processing Systems
Scott Yih, Microsoft: Machine Learning for Language Processing
Luke Zettlemoyer, University of Washington: Syntax and Parsing
Bowen Zhou, IBM: Machine Translation
Jerry Zhu, University of Wisconsin: Document Categorization/Topic Clustering

Local Arrangements:

Chair:

Priscilla Rasmussen, ACL Business Office, acl-AT-aclweb.org

Advisory committee:

Sabine Bergler, Concordia University
Guy Lapalme, Université de Montréal

Workshops Co-Chairs:

Colin Cherry, National Research Council of Canada
Mona Diab, Columbia University

Tutorials Co-Chairs:

Jacob Eisenstein, CMU
Radu Florian, IBM T.J. Watson Research Center

Demos Co-Chairs:

Aria Haghighi, Prismatic
Yaser Al-Onaizan, IBM T.J. Watson Research Center

Student Workshop:

Co-chairs:

Rivka Levitan, Columbia University
Myle Ott, Cornell University

Faculty Advisors:

Roger Levy, UCSD
Ani Nenkova, University of Pennsylvania

Publications:

Nizar Habash, Columbia University
William Schuler, OSU

Publicity:

Smaranda Muresan, Rutgers University

Exhibits:

Joel Tetreault, Education Testing Services

Webmaster:

Dirk Hovy, USC/ISI

Americas Sponsorship Co-chairs:

Michael Gamon, Microsoft Research
Patrick Pantel, Microsoft Research

Program Committee Members:

Hua Ai
Gregory Aist
Cem Akkaya
Afra Alishahi
Cecilia O. Alm
Giambattista Amati
Ion Androutsopoulos
David Andrzejewski
Mark Arehart
Victoria Arranz
Javier Artiles
Ron Artstein
Abhishek Arun
Javed Aslam
Giuseppe Attardi
Necip Fazil Ayan
Michiel Bacchiani
Timothy Baldwin
Krisztian Balog
Eva Banik
Michele Banko
Ken Barker
Roberto Bayardo
Nick Belkin
Anja Belz
Emily Bender
Michael Bendersky
Paul Bennett
Edward Benson
Sabine Bergler
Shane Bergsma
Raffaella Bernardi
Steven Bethard
Rahul Bhagat
Alan W. Black
Roi Blanco
John Blitzer
Branimir Boguraev
Dan Bohus
Johan Bos
Alexandre Bouchard-Cote
Gosse Bouma
Jordan Boyd-Graber
Kristy Boyer
Thorsten Brants
Ulf Brefeld
Fabio Brugnara
Sabine Buchholz
Bill Byrne
Aoife Cahill
Clarie Cardie
Andrew Carlson
Marine Carpuat
John Carroll
Francisco Casacuberta
Diamantino Caseiro
Tommaso Caselli
Mauro Cettolo
Joyce Chai
Yllias Chali
Ming-Wei Chang
Pi-Chuan Chang
Eugene Charniak
Ciprian Chelba
Boxing Chen
John Chen
Zheng Chen
Wenliang Chen
Colin Cherry
David Chiang
Yejin Choi
Christos Christodoulopoulos
Wei Chu
Philipp Cimiano
Stephen Clark
Alex Clark
Martin Cmejrek
Shay Cohen
Ronan Collobert
Sherri Condon
Gao Cong
Michael Connor
Paul Cook
Mark Core
Stephen Cox
Mark Craven
Danilo Croce
Aron Culotta
Walter Daelemans
Geraldine Damnati
Hoa Trang Dang
Dipanjan Das
Sajib Dasgupta
Hal Daume III
Adria de Gispert
Renato De Mori
Rodolfo Delmonte
Dina Demner-Fushman
Steve DeNeefe
John DeNero
Hongbo Deng
Nicholas Diakopoulos
Laura Dietz
Chrysanne DiMarco
Xiaowen Ding
Denise DiPersio
Doug Downey
Markus Dreyer
Kevin Duh
Chris Dyer
Marc Dymetman
Myroslava Dzikovska
Jens Edlund
Koji Eguchi
Jacob Eisenstein
Charles Elkan
Jonathan Elsas
Micha Elsner
Ahmad Emami
Hakan Erdogan
Hui Fang
Faisal Farooq
Afsaneh Fazly
Anna Feldman
Donghui Feng
Elena Filatova

Katja Filippova
Tim Finin
Jenny Finkel
Dan Flickinger
Radu Florian
Kate Forbes-Riley
Jennifer Foster
Francesca Frontini
Michel Galley
Kuzman Ganchev
Kavita Ganesan
Jianfeng Gao
Qin Gao
Albert Gatt
Niyu Ge
Dmitriy Genzel
Kallirroi Georgila
Panayiotis Georgiou
Matthew Gerber
Shalini Ghosh
Arnab Ghoshal
Daniel Gildea
Kevin Gimpel
Roxana Girju
Alfio Gliozzo
Amir Globerson
Vibhav Gogate
Yoav Goldberg
Sharon Goldwater
Carlos Gomez Rodriguez
Julio Gonzalo
Joao Graca
Agustin Gravano
Nancy Green
Gregory Grefenstette
Ralph Grishman
Markus Guhe
Iryna Gurevych
Nizar Habash
Gholamreza Haffari
Dilek Hakkani-Tur
Keith Hall
David Hall
Olivier Hamon
Sandra Harabagin
Lisa Harper
Helen Hastie
Timothy J. Hazen
Xiaodong He
Marti Hearst
Jeffrey Heinz
Aurelie Herbelot
Sanjika Hewavitharana
Silja Hildebrand
Graeme Hirst
Barbora Hladka
Hieu Hoang
Julia Hockenmaier
Minlie Huang
Zhongqiang Huang
Jimmy Huang
Yang Hui
Nancy Ide
Amy Isard
Abraham Ittycheriah
Emily Jamison
Jing Jiang
Richard Johansson
Howard Johnson
Mark Johnson
Kristiina Jokinen
Doug Jones
Pamela Jordan
Shafiq Joty
Nobuhiro Kaji
Hiroshi Kanayama
Daisuke Kawahara
Anna Kazantseva
Alistair Kennedy
Joseph Keshet
Tracy King
Brian Kingsbury
Katrin Kirchhoff
Dietrich Klakow
Alexandre Klementiev
Philipp Koehn
Rob Koeling
Stanley Kok
Alexander Koller
Grzegorz Kondrak
Terry Koo
Zornitsa Kozareva
Taku Kudoh
Sandra Kuebler
Marco Kuhlmann
Roland Kuhn
Seth Kulick
Shankar Kumar
Oren Kurland
Philippe Langlais
Guy Lapalme
Mirella Lapata
Alon Lavie
Matt Lease
John Lehmann
Alessandro Lenci
James Lester
Gregor Leusch
Anton Leuski
Haibo Li
Shoushan Li
Hang Li
Zhifei Li
Fangtao Li
Mu Li
Fennie Liang
Percy Liang
Elizabeth Liddy
Ding Liu
Zhiyuan Liu
Qun Liu
Yang Liu
Yan Liu
Eduardo Lleida Solano
Oier Lopez de Lacalle
Annie Louis
Pengfei Lu
Bin Lu
Yue Lu
Klaus Macherey
Wolfgang Macherey
Nitin Madnani
Andreas Maletti
Arindam Mandal
Gideon Mann
Christopher Manning
Daniel Marcu
James Martin

Andre Martins
Yuval Marton
Sameer Maskey
Spyros Matsoukas
Takuya Matsuzaki
Mike Maxwell
Jon May
James Mayfield
Diana Maynard
Diana McCarthy
Ryan McDonald
Paul McNamee
Dan Melamed
Chris Mellish
Arul Menezes
Donald Metzler
Haitao Mi
Rada Mihalcea
Keith J. Miller
Andriy Mnih
Saif Mohammad
Bob Moore
Alessandro Moschitti
Dragos Munteanu
Smaranda Muresan
Gabriel Murray
Meenakshi Nagarajan
Tetsuji Nakagawa
Roberto Navigli
Mark-Jan Nederhof
Ani Nenkova
Alena Neviarouskaya
Hwee Tou Ng
Alexandru Niculescu-Mizil
Jian-Yun Nie
Joakim Nivre
Stephan Oepen
Constantin Orasan
Beatrice Oshika
Christopher Pal
Sinno Jialin Pan
Bo Pang
Soo-Min Pantel
Cecile Paris
Chris Parisien
S Parthasarathy
Marius Pasca
Siddharth Patwardhan
Michael Paul
Adam Pauls
Anselmo Penas
Marco Pennacchiotti
Slav Petrov
Fabio Pianesi
Emily Pitler
Paul Piwek
Thierry Poibeau
Joe Polifroni
Hoifung Poon
Martin Popel
Adrian Popescu
Maja Popovic
Matt Post
Sameer Pradhan
John Prager
Mark Przybocki
Stephen Pulman
Matthew Purver
Yanjun Qi
Silvia Quarteroni
Chris Quirk
Dragomir Radev
Piyush Rai
Owen Rambow
Delip Rao
Ari Rappoport
Lev-Arie Ratinov
Sujith Ravi
Sravana Reddy
Ines Rehbein
Ehud Reiter
Jason Riggle
Laura Rimell
Alan Ritter
Antonio Roque
Carolyn Rose
Andrew Rosenberg
Afshin Rostamizadeh
Dan Roth
Marta Ruiz
Josef Ruppenhofer
Irene Russo
Kenji Sagae
Alicia Sagae
Horacio Saggion
Murat Saraclar
Anoop Sarkar
Manabu Sassano
Hassan Sawaf
Frank Schilder
David Schlangen
Sabine Schulte im Walde
Holger Schwenk
Donia Scott
Frédérique Segond
Satoshi Sekine
Hendra Setiawan
Burr Settles
Wade Shen
Shuming Shi
Michel Simard
Vikas Sindhwani
Kevin Small
David Smith
Ronnie Smith
Benjamin Snyder
Morgan Sonderegger
Lucia Specia
Caroline Sporleder
Matthew Stone
Veselin Stoyanov
Carlo Strapparava
Stephanie Strassel
Kristina Striegnitz
Michael Strube
Jian Su
L Venkata Subramaniam
Amarnag Subramanya
Richard Sutcliffe
Charles Sutton
Hiroya Takamura
Partha Talukdar
Joerg Tiedemann
Christoph Tillmann
Ivan Titov
Noriko Tomuro
Sara Tonelli
Audrey Tong

Cigdem Toprak
Kristina Toutanova
Vivian Tsang
Oren Tsur
Yoshimasa Tsuruoka
Gokhan Tur
Joseph Turian
Oscar Täckström
Jacob Uszkoreit
Lonneke van der Plas
Ielka van der Sluis
Gertjan van Noord
Lucy Vanderwende
Ashish Venugopal
Guido Vetere
David Vilar
Aline Villavicencio
Ellen Voorhees
Xiaojun Wan
Stephen Wan
Lidan Wang
Chong Wang
Haifeng Wang
Nigel Ward
Taro Watanabe
Bonnie Webber
Fuliang Weng
Michael White
Richard Wicentowski
Jason Williams
Theresa Wilson
Guillaume Wisniewski
Kamfai Wong
Frank Wood
Dekai Wu
Yunqing Xia
Bing Xiang
Deyi Xiong
Peng Xu
Nianwen Xue
Fan Yang
Muyun Yang
Tianbao Yang
Roman Yangarber
Alex Yates
Ainur Yessenalina
Yisong Yue
Murawaki Yugo
François Yvon
Fabio Massimo Zanzotto
Beñat Zapirain
Richard Zens
Zhongwu Zhai
Chengxiang Zhai
Hao Zhang
Ying Zhang
Min Zhang
Lei Zhang
Yi Zhang
Tiejun Zhao
Bing Zhao
Jing Zheng
Zhi-Hua Zhou
Dengyong Zhou
Guodong Zhou
Ming Zhou
Jerry Zhu
Xiaodan Zhu
Imed Zitouni
Chengqing Zong
Geoffrey Zweig
Diarmuid Ó Séaghdha
Lilja Øvrelid

Table of Contents

Multiple Narrative Disentanglement: Unraveling Infinite Jest
Byron Wallace 1

Acoustic-Prosodic Entrainment and Social Behavior
Rivka Levitan, Agustin Gravano, Laura Willson, Stefan Benus, Julia Hirschberg and Ani Nenkova
11

Identifying High-Level Organizational Elements in Argumentative Discourse
Nitin Madnani, Michael Heilman, Joel Tetreault and Martin Chodorow 20

Fast Inference in Phrase Extraction Models with Belief Propagation
David Burkett and Dan Klein 29

Continuous Space Translation Models with Neural Networks
Hai-Son Le, Alexandre Allauzen and François Yvon 39

Machine Translation of Arabic Dialects
Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F. Zaidan and Chris Callison-Burch 49

Entity Clustering Across Languages
Spence Green, Nicholas Andrews, Matthew R. Gormley, Mark Dredze and Christopher D. Manning 60

Multi-Event Extraction Guided by Global Constraints
Roi Reichart and Regina Barzilay 70

Reference Scope Identification in Citing Sentences
Amjad Abu Jbara and Dragomir Radev 80

Intrinsic and Extrinsic Evaluation of an Automatic User Disengagement Detector for an Uncertainty-Adaptive Spoken Dialogue System
Kate Forbes-Riley, Diane Litman, Heather Friedberg and Joanna Drummond 91

Exploring Content Features for Automated Speech Scoring
Shasha Xie, Keelan Evanini and Klaus Zechner 103

Hello, Who is Calling?: Can Words Reveal the Social Nature of Conversations?
Anthony Stark, Izhak Shafran and Jeffrey Kaye 112

Minimum-Risk Training of Approximate CRF-Based NLP Systems
Veselin Stoyanov and Jason Eisner 120

Unsupervised Learning on an Approximate Corpus
Jason Smith and Jason Eisner 131

Structured Perceptron with Inexact Search
Liang Huang, Suphan Fayong and Yang Guo 142

Segmentation Similarity and Agreement
Chris Fournier and Diana Inkpen 152

HyTER: Meaning-Equivalent Semantics for Translation Evaluation
Markus Dreyer and Daniel Marcu 162

Apples to Oranges: Evaluating Image Annotations from Natural Language Processing Systems
Rebecca Mason and Eugene Charniak 172

Re-examining Machine Translation Metrics for Paraphrase Identification
Nitin Madnani, Joel Tetreault and Martin Chodorow 182

A Dependency Treebank of Classical Chinese Poems
John Lee and Yin Hei Kong 191

Towards Effective Tutorial Feedback for Explanation Questions: A Dataset and Baselines
Myroslava O. Dzikovska, Rodney D. Nielsen and Chris Brew 200

Topical Segmentation: a Study of Human Performance and a New Measure of Quality.
Anna Kazantseva and Stan Szpakowicz 211

Structured Ramp Loss Minimization for Machine Translation
Kevin Gimpel and Noah A. Smith 221

Implicitly Intersecting Weighted Automata using Dual Decomposition
Michael J. Paul and Jason Eisner 232

Transliteration Mining Using Large Training and Test Sets
Ali El-Kahki, Kareem Darwish, Mohamed Abdul-Wahab and Ahmed Taei 243

Optimized Online Rank Learning for Machine Translation
Taro Watanabe 253

Every sensible extended top-down tree transducer is a multi bottom-up tree transducer
Andreas Maletti 263

NOMIT: Automatic Titling by Nominalizing
Cédric Lopez, Violaine Prince and Mathieu Roche 274

Correcting Comma Errors in Learner Essays, and Restoring Commas in Newswire Text
Ross Israel, Joel Tetreault and Martin Chodorow 284

The Challenges of Parsing Chinese with Combinatory Categorial Grammar
Daniel Tse and James R. Curran 295

Using Supertags and Encoded Annotation Principles for Improved Dependency to Phrase Structure Conversion
Seth Kulick, Ann Bies and Justin Mott 305

Getting More from Morphology in Multilingual Dependency Parsing
Matt Hohensee and Emily M. Bender 315

Stylometric Analysis of Scientific Articles
Shane Bergsma, Matt Post and David Yarowsky 327

Using paraphrases for improving first story detection in news and Twitter
Sasa Petrovic, Miles Osborne and Victor Lavrenko 338

Insertion and Deletion Models for Statistical Machine Translation
Matthias Huck and Hermann Ney 347

TransAhead: A Computer-Assisted Translation and Writing Tool
Chung-chi Huang, Ping-che Yang, Keh-jiann Chen and Jason S. Chang 352

Correction Detection and Error Type Selection as an ESL Educational Aid
Ben Swanson and Elif Yamangil 357

Getting More from Segmentation Evaluation
Martin Scaiano and Diana Inkpen 362

G2P Conversion of Proper Names Using Word Origin Information
Sonjia Waxmonsky and Sravana Reddy 367

Evaluating a Morphological Analyser of Inuktitut
Jeremy Nicholson, Trevor Cohn and Timothy Baldwin 372

Intra-Speaker Topic Modeling for Improved Multi-Party Meeting Summarization with Integrated Random Walk
Yun-Nung Chen and Florian Metze 377

Towards Using EEG to Improve ASR Accuracy
Yun-Nung Chen, Kai-Min Chang and Jack Mostow 382

A Comparative Investigation of Morphological Language Modeling for the Languages of the European Union
Thomas Mueller, Hinrich Schuetze and Helmut Schmid 386

Leveraging supplemental representations for sequential transduction
Aditya Bhargava and Grzegorz Kondrak 396

A Hierarchical Dirichlet Process Model for Joint Part-of-Speech and Morphology Induction
Kairit Sirts and Tanel Alumäe 407

Encouraging Consistent Translation Choices
Ferhan Ture, Douglas W. Oard and Philip Resnik 417

Batch Tuning Strategies for Statistical Machine Translation
Colin Cherry and George Foster 427

Real-time Incremental Speech-to-Speech Translation of Dialogs
Srinivas Bangalore, Vivek Kumar Rangarajan Sridhar, Prakash Kolan, Ladan Golipour and Aura Jimenez 437

Parsing Time: Learning to Interpret Time Expressions
Gabor Angeli, Christopher Manning and Daniel Jurafsky 446

Fine-Grained Focus for Pinpointing Positive Implicit Meaning from Negated Statements
Eduardo Blanco and Dan Moldovan 456

Taxonomy Induction Using Hierarchical Random Graphs
Trevor Fountain and Mirella Lapata 466

Cross-lingual Word Clusters for Direct Transfer of Linguistic Structure
Oscar Täckström, Ryan McDonald and Jakob Uszkoreit 477

Training Dependency Parser Using Light Feedback
Avihai Mejer and Koby Crammer 488

Vine Pruning for Efficient Multi-Pass Dependency Parsing
Alexander Rush and Slav Petrov 498

Active Learning for Coreference Resolution
Florian Laws, Florian Heimerl and Hinrich Schütze 508

Space Efficiencies in Discourse Modeling via Conditional Random Sampling
Brian Kjersten and Benjamin Van Durme 513

Predicting Overt Display of Power in Written Dialogs
Vinodkumar Prabhakaran, Owen Rambow and Mona Diab 518

Co-reference via Pointing and Haptics in Multi-Modal Dialogues
Lin Chen and Barbara Di Eugenio 523

Trait-Based Hypothesis Selection For Machine Translation
Jacob Devlin and Spyros Matsoukas 528

Improved Reordering for Shallow-n Grammar based Hierarchical Phrase-based Translation
Baskaran Sankaran and Anoop Sarkar 533

Automatic Parallel Fragment Extraction from Noisy Data
Jason Riesa and Daniel Marcu 538

Tuning as Linear Regression
Marzieh Bazrafshan, Tagyoung Chung and Daniel Gildea 543

Ranking-based readability assessment for early primary children's literature
Yi Ma, Eric Fosler-Lussier and Robert Lofthus 548

How Text Segmentation Algorithms Gain from Topic Models
Martin Riedl and Chris Biemann 553

Identifying Comparable Corpora Using LDA
Judita Preiss 558

Behavioral Factors in Interactive Training of Text Classifiers
Burr Settles and Xiaojin Zhu 563

Better Evaluation for Grammatical Error Correction
Daniel Dahlmeier and Hwee Tou Ng 568

Are You Sure? Confidence in Prediction of Dependency Tree Edges
Avihai Mejer and Koby Crammer 573

Concavity and Initialization for Unsupervised Dependency Parsing
Kevin Gimpel and Noah A. Smith 577

Multimodal Grammar Implementation
Katya Alahverdzhieva, Dan Flickinger and Alex Lascarides 582

Portable Features for Classifying Emotional Text
Saif Mohammad 587

Stance Classification using Dialogic Properties of Persuasion
Marilyn Walker, Pranav Anand, Rob Abbott and Ricky Grant 592

Context-Enhanced Citation Sentiment Detection
Awais Athar and Simone Teufel 597

Predicting Responses to Microblog Posts
Yoav Artzi, Patrick Pantel and Michael Gamon 602

The Intelius Nickname Collection: Quantitative Analyses from Billions of Public Records
Vitor Carvalho, Yigit Kiran and Andrew Borthwick 607

A comparison of models of word meaning in context
Georgiana Dinu, Stefan Thater and Soeren Laue 611

Measuring Word Relatedness Using Heterogeneous Vector Space Models
Wen-tau Yih and Vahed Qazvinian 616

Expectations of Word Sense in Parallel Corpora
Xuchen Yao, Benjamin Van Durme and Chris Callison-Burch 621

Why Not Grab a Free Lunch? Mining Large Corpora for Parallel Sentences to Improve Translation Modeling
Ferhan Ture and Jimmy Lin 626

Summarization of Historical Articles Using Temporal Event Clustering
James Gung and Jugal Kalita 631

Comparing HMMs and Bayesian Networks for Surface Realisation
Nina Dethlefs and Heriberto Cuayahuitl 636

On The Feasibility of Open Domain Referring Expression Generation Using Large Scale Folksonomies
Fabián Pacheco, Pablo Duboue and Martín Domínguez 641

Structured Event Retrieval over Microblog Archives
Donald Metzler, Congxing Cai and Eduard Hovy 646

Learning from Bullying Traces in Social Media
Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu and Amy Bellmore 656

Grammatical structures for word-level sentiment detection
Asad Sayeed, Jordan Boyd-Graber, Bryan Rusk and Amy Weinberg 667

Graph-Based Lexicon Expansion with Sparsity-Inducing Penalties
Dipanjan Das and Noah A. Smith 677

Unified Expectation Maximization
Rajhans Samdani, Ming-Wei Chang and Dan Roth 688

Low-Dimensional Discriminative Reranking
Jagadeesh Jagarlamudi and Hal Daume III 699

Autonomous Self-Assessment of Autocorrections: Exploring Text Message Dialogues
Tyler Baldwin and Joyce Chai 710

Translation-Based Projection for Multilingual Coreference Resolution
Altaf Rahman and Vincent Ng 720

Exploring Semi-Supervised Coreference Resolution of Medical Concepts using Semantic and Temporal Features
Preethi Raghavan, Eric Fosler-Lussier and Albert Lai 731

Mind the Gap: Learning to Choose Gaps for Question Generation
Lee Becker, Sumit Basu and Lucy Vanderwende 742

Unsupervised Concept-to-text Generation with Hypergraphs
Ioannis Konstas and Mirella Lapata 752

Detecting Visual Text
Jesse Dodge, Amit Goyal, Xufeng Han, Alyssa Mensch, Margaret Mitchell, Karl Stratos, Kota Yamaguchi, Yejin Choi, Hal Daume III, Alex Berg and Tamara Berg 762

Unsupervised Translation Sense Clustering
Mohit Bansal, John DeNero and Dekang Lin 773

Shared Components Topic Models
Matthew R. Gormley, Mark Dredze, Benjamin Van Durme and Jason Eisner 783

Textual Predictors of Bill Survival in Congressional Committees
Tae Yano, Noah A. Smith and John D. Wilkerson 793

Conference Program

Monday, June 4, 2012

(7:30-5:00) Registration

(7:30-9:00) Breakfast

(9:00-9:15) Welcoming remarks

(9:15-10:30) Keynote: Eduard Hovy – "A New Semantics: Merging Propositional and Distributional Information"

(10:30-11:00) Coffee Break

Session Mon-1E: (11:00-12:30) Discourse, Dialog, and Pragmatics I

11:00–11:30 *Multiple Narrative Disentanglement: Unraveling Infinite Jest*
Byron Wallace

11:30–12:00 *Acoustic-Prosodic Entrainment and Social Behavior*
Rivka Levitan, Agustin Gravano, Laura Willson, Stefan Benus, Julia Hirschberg and Ani Nenkova

12:00–12:30 *Identifying High-Level Organizational Elements in Argumentative Discourse*
Nitin Madnani, Michael Heilman, Joel Tetreault and Martin Chodorow

Session Mon-1W: (11:00-12:30) Machine Translation I

11:00–11:30 *Fast Inference in Phrase Extraction Models with Belief Propagation*
David Burkett and Dan Klein

11:30–12:00 *Continuous Space Translation Models with Neural Networks*
Hai-Son Le, Alexandre Allauzen and François Yvon

12:00–12:30 *Machine Translation of Arabic Dialects*
Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F. Zaidan and Chris Callison-Burch

Monday, June 4, 2012 (continued)

Session Mon-1D: (11:00-12:30) Information Extraction

11:00–11:30 *Entity Clustering Across Languages*
Spence Green, Nicholas Andrews, Matthew R. Gormley, Mark Dredze and Christopher D. Manning

11:30–12:00 *Multi-Event Extraction Guided by Global Constraints*
Roi Reichart and Regina Barzilay

12:00–12:30 *Reference Scope Identification in Citing Sentences*
Amjad Abu Jbara and Dragomir Radev

(12:30-1:15) Student lunch sponsored by IBM and the Student Research Workshop (students only)

(1:15-2:15) SRW Panel: Reviewing Practices (open to all)

Session Mon-2E: (2:30-4:00) Spoken Language Processing

2:30–3:00 *Intrinsic and Extrinsic Evaluation of an Automatic User Disengagement Detector for an Uncertainty-Adaptive Spoken Dialogue System*
Kate Forbes-Riley, Diane Litman, Heather Friedberg and Joanna Drummond

3:00–3:30 *Exploring Content Features for Automated Speech Scoring*
Shasha Xie, Keelan Evanini and Klaus Zechner

3:30–4:00 *Hello, Who is Calling?: Can Words Reveal the Social Nature of Conversations?*
Anthony Stark, Izhak Shafran and Jeffrey Kaye

Monday, June 4, 2012 (continued)

Session Mon-2W: (2:30-4:00) Machine Learning I

2:30–3:00 *Minimum-Risk Training of Approximate CRF-Based NLP Systems*
Veselin Stoyanov and Jason Eisner

3:00–3:30 *Unsupervised Learning on an Approximate Corpus*
Jason Smith and Jason Eisner

3:30–4:00 *Structured Perceptron with Inexact Search*
Liang Huang, Suphan Fayong and Yang Guo

Session Mon-2D: (2:30-4:00) Language Resources and Evaluations

2:30–3:00 *Segmentation Similarity and Agreement*
Chris Fournier and Diana Inkpen

3:00–3:30 *HyTER: Meaning-Equivalent Semantics for Translation Evaluation*
Markus Dreyer and Daniel Marcu

3:30–4:00 *Apples to Oranges: Evaluating Image Annotations from Natural Language Processing Systems*
Rebecca Mason and Eugene Charniak

(4:00-4:30) Coffee Break

(4:30-5:30) Posters and Demos: One-Minute Madness

Monday, June 4, 2012 (continued)

(6:00-9:00) Poster and Demo Session (with Buffet Dinner)

Session Mon-P: Posters: Full Papers

Re-examining Machine Translation Metrics for Paraphrase Identification
Nitin Madnani, Joel Tetreault and Martin Chodorow

A Dependency Treebank of Classical Chinese Poems
John Lee and Yin Hei Kong

Towards Effective Tutorial Feedback for Explanation Questions: A Dataset and Baselines
Myroslava O. Dzikovska, Rodney D. Nielsen and Chris Brew

Topical Segmentation: a Study of Human Performance and a New Measure of Quality.
Anna Kazantseva and Stan Szpakowicz

Structured Ramp Loss Minimization for Machine Translation
Kevin Gimpel and Noah A. Smith

Implicitly Intersecting Weighted Automata using Dual Decomposition
Michael J. Paul and Jason Eisner

Transliteration Mining Using Large Training and Test Sets
Ali El-Kahki, Kareem Darwish, Mohamed Abdul-Wahab and Ahmed Taei

Optimized Online Rank Learning for Machine Translation
Taro Watanabe

Every sensible extended top-down tree transducer is a multi bottom-up tree transducer
Andreas Maletti

NOMIT: Automatic Titling by Nominalizing
Cédric Lopez, Violaine Prince and Mathieu Roche

Correcting Comma Errors in Learner Essays, and Restoring Commas in Newswire Text
Ross Israel, Joel Tetreault and Martin Chodorow

Using Supertags and Encoded Annotation Principles for Improved Dependency to Phrase Structure Conversion
Seth Kulick, Ann Bies and Justin Mott

Stylometric Analysis of Scientific Articles
Shane Bergsma, Matt Post and David Yarowsky

Using paraphrases for improving first story detection in news and Twitter
Sasa Petrovic, Miles Osborne and Victor Lavrenko

Session Mon-PS: Posters: Short Papers

TransAhead: A Computer-Assisted Translation and Writing Tool
Chung-chi Huang, Ping-che Yang, Keh-jiann Chen and Jason S. Chang

Correction Detection and Error Type Selection as an ESL Educational Aid
Ben Swanson and Elif Yamangil

Getting More from Segmentation Evaluation
Martin Scaiano and Diana Inkpen

G2P Conversion of Proper Names Using Word Origin Information
Sonjia Waxmonsky and Sravana Reddy

Evaluating a Morphological Analyser of Inuktitut
Jeremy Nicholson, Trevor Cohn and Timothy Baldwin

Intra-Speaker Topic Modeling for Improved Multi-Party Meeting Summarization with Integrated Random Walk
Yun-Nung Chen and Florian Metze

Towards Using EEG to Improve ASR Accuracy
Yun-Nung Chen, Kai-Min Chang and Jack Mostow

Posters: Student Research Workshop

Domain-Specific Semantic Relatedness From Wikipedia: Can A Course Be Transferred?
Beibei Yang and Jesse M. Heines

Automatic Animacy Classification
Samuel Bowman and Harshit Chopra

Finding the Right Supervisor: Expert-Finding in a University Domain
Fawaz Alarfaj, Udo Kruschwitz, David Hunter and Chris Fox

Indexing Google 1T for low-turnaround wildcarded frequency queries
Steinar Kaldager

Automatic Humor Classification on Twitter
Yishay Raz

Beauty Before Age? Applying Subjectivity to Automatic English Adjective Ordering
Felix Hill

Automatic Metrics for Genre-specific Text Quality
Annie Louis

A Weighting Scheme for Open Information Extraction
Yuval Merhav

Choosing an Evaluation Metric for Parser Design
Woodley Packard

Using Ontology-based Approaches to Representing Speech Transcripts for Automated Speech Scoring
Miao Chen

Deep Unsupervised Feature Learning for Natural Language Processing
Stephan Gouws

Unified Extraction of Health Condition Descriptions
Ivelina Nikolova

Posters: Demo

DeSoCoRe: Detecting Source Code Re-Use across Programming Languages
Enrique Flores, Alberto Barrón-Cedeño, Paolo Rosso and Lidia Moreno

A Graphical User Interface for Feature-Based Opinion Mining
Pedro Paulo Balage Filho, Caroline Brun and Gilbert Rondeau

Navigating Large Comment Threads with CoFi
Christine Doran, Guido Zarrella and John C. Henderson

SurfShop: combing a product ontology with topic model results for online window-shopping.
Zofia Stankiewicz and Satoshi Sekine

An Interactive Humanoid Robot Exhibiting Flexible Sub-Dialogues
Heriberto Cuayáhuitl and Ivana Kruijff-Korbayová

MSR SPLAT, a language analysis toolkit
Chris Quirk, Pallavi Choudhury, Jianfeng Gao, Hisami Suzuki, Kristina Toutanova, Michael Gamon, Wen-tau Yih, Colin Cherry and Lucy +Vanderwende

Incremental Speech Understanding in a Multi-Party Virtual Human Dialogue System
David DeVault and David Traum

A Robust Shallow Temporal Reasoning System
Ran Zhao, Quang Do and Dan Roth

AttitudeMiner: Mining Attitude from Online Discussions
Amjad Abu-Jbara, Ahmed Hassan and Dragomir Radev

Tuesday, June 5, 2012

(7:30-5:00) Registration

(7:30-9:00) Breakfast

Session Tue-3: (9:00-10:30) Best Paper Awards Session

9:10–9:30 *Trait-Based Hypothesis Selection For Machine Translation*
Jacob Devlin and Spyros Matsoukas

9:30–10:00 *Cross-lingual Word Clusters for Direct Transfer of Linguistic Structure*
Oscar Täckström, Ryan McDonald and Jakob Uszkoreit

10:00–10:30 *Vine Pruning for Efficient Multi-Pass Dependency Parsing*
Alexander Rush and Slav Petrov

(10:30-11:00) Coffee Break

Session Tue-4E: (11:00-12:30) Phonology and Morphology

11:00–11:30 *A Comparative Investigation of Morphological Language Modeling for the Languages of the European Union*
Thomas Mueller, Hinrich Schuetze and Helmut Schmid

11:30–12:00 *Leveraging supplemental representations for sequential transduction*
Aditya Bhargava and Grzegorz Kondrak

12:00–12:30 *A Hierarchical Dirichlet Process Model for Joint Part-of-Speech and Morphology Induction*
Kairit Sirts and Tanel Alumäe

Tuesday, June 5, 2012 (continued)

Session Tue-4C: (11:00-12:30) Machine Translation II

11:00–11:30 *Encouraging Consistent Translation Choices*
Ferhan Ture, Douglas W. Oard and Philip Resnik

11:30–12:00 *Batch Tuning Strategies for Statistical Machine Translation*
Colin Cherry and George Foster

12:00–12:30 *Real-time Incremental Speech-to-Speech Translation of Dialogs*
Srinivas Bangalore, Vivek Kumar Rangarajan Sridhar, Prakash Kolan, Ladan Golipour and Aura Jimenez

Session Tue-4W: (11:00-12:30) Semantics I

11:00–11:30 *Parsing Time: Learning to Interpret Time Expressions*
Gabor Angeli, Christopher Manning and Daniel Jurafsky

11:30–12:00 *Fine-Grained Focus for Pinpointing Positive Implicit Meaning from Negated Statements*
Eduardo Blanco and Dan Moldovan

12:00–12:30 *Taxonomy Induction Using Hierarchical Random Graphs*
Trevor Fountain and Mirella Lapata

Session Tue-4D: (11:00-12:30) Syntax and Parsing

11:00–11:30 *Getting More from Morphology in Multilingual Dependency Parsing*
Matt Hohensee and Emily M. Bender

11:30–12:00 *Training Dependency Parser Using Light Feedback*
Avihai Mejer and Koby Crammer

12:00–12:30 *The Challenges of Parsing Chinese with Combinatory Categorial Grammar*
Daniel Tse and James R. Curran

Tuesday, June 5, 2012 (continued)

(12:30-2:00) Lunch

(2:00-3:30) NLP Idol: Plucked from Obscurity

(3:30-4:00) Coffee Break

Session Tue-5E: (4:00-5:20) Short papers: Discourse

4:00–4:20 *Active Learning for Coreference Resolution*
Florian Laws, Florian Heimerl and Hinrich Schütze

4:20–4:40 *Space Efficiencies in Discourse Modeling via Conditional Random Sampling*
Brian Kjersten and Benjamin Van Durme

4:40–5:00 *Predicting Overt Display of Power in Written Dialogs*
Vinodkumar Prabhakaran, Owen Rambow and Mona Diab

5:00–5:20 *Co-reference via Pointing and Haptics in Multi-Modal Dialogues*
Lin Chen and Barbara Di Eugenio

Session Tue-5C: (4:00-5:20) Short papers: MT

4:00–4:20 *Insertion and Deletion Models for Statistical Machine Translation*
Matthias Huck and Hermann Ney

4:20–4:40 *Improved Reordering for Shallow-n Grammar based Hierarchical Phrase-based Translation*
Baskaran Sankaran and Anoop Sarkar

4:40–5:00 *Automatic Parallel Fragment Extraction from Noisy Data*
Jason Riesa and Daniel Marcu

5:00–5:20 *Tuning as Linear Regression*
Marzieh Bazrafshan, Tagyoung Chung and Daniel Gildea

Tuesday, June 5, 2012 (continued)

Session Tue-5W: (4:00-5:20) Short papers: Document Categorization and Topic Modeling

4:00–4:20 *Ranking-based readability assessment for early primary children's literature*
Yi Ma, Eric Fosler-Lussier and Robert Lofthus

4:20–4:40 *How Text Segmentation Algorithms Gain from Topic Models*
Martin Riedl and Chris Biemann

4:40–5:00 *Identifying Comparable Corpora Using LDA*
Judita Preiss

5:00–5:20 *Behavioral Factors in Interactive Training of Text Classifiers*
Burr Settles and Xiaojin Zhu

Session Tue-5D: (4:00-5:20) Short papers: Syntax

4:00–4:20 *Better Evaluation for Grammatical Error Correction*
Daniel Dahlmeier and Hwee Tou Ng

4:20–4:40 *Are You Sure? Confidence in Prediction of Dependency Tree Edges*
Avihai Mejer and Koby Crammer

4:40–5:00 *Concavity and Initialization for Unsupervised Dependency Parsing*
Kevin Gimpel and Noah A. Smith

5:00–5:20 *Multimodal Grammar Implementation*
Katya Alahverdzhieva, Dan Flickinger and Alex Lascarides

Tuesday, June 5, 2012 (continued)

(7:00) Banquet at Le Windsor Ballroom

Wednesday, June 6, 2012

(7:30-5:00) Registration

(7:30-9:00) Breakfast

(9:00-10:15) Keynote: James W. Pennebaker – "A, is, I, and, the: How our smallest words reveal the most about who we Are"

(10:15-10:40) Coffee Break

Session Wed-6E: (10:40-12:00) Short papers: Sentiment and Social Media

10:40–11:00 *Portable Features for Classifying Emotional Text*
Saif Mohammad

11:00–11:20 *Stance Classification using Dialogic Properties of Persuasion*
Marilyn Walker, Pranav Anand, Rob Abbott and Ricky Grant

11:20–11:40 *Context-Enhanced Citation Sentiment Detection*
Awais Athar and Simone Teufel

11:40–12:00 *Predicting Responses to Microblog Posts*
Yoav Artzi, Patrick Pantel and Michael Gamon

Wednesday, June 6, 2012 (continued)

Session Wed-6C: (10:40-12:00) Short papers: Semantics

10:40–11:00 *The Intelius Nickname Collection: Quantitative Analyses from Billions of Public Records*
Vitor Carvalho, Yigit Kiran and Andrew Borthwick

11:00–11:20 *A comparison of models of word meaning in context*
Georgiana Dinu, Stefan Thater and Soeren Laue

11:20–11:40 *Measuring Word Relatedness Using Heterogeneous Vector Space Models*
Wen-tau Yih and Vahed Qazvinian

11:40–12:00 *Expectations of Word Sense in Parallel Corpora*
Xuchen Yao, Benjamin Van Durme and Chris Callison-Burch

Session Wed-6W: (10:40-12:00) Short papers: Summarization

10:40–11:00 *Why Not Grab a Free Lunch? Mining Large Corpora for Parallel Sentences to Improve Translation Modeling*
Ferhan Ture and Jimmy Lin

11:00–11:20 *Summarization of Historical Articles Using Temporal Event Clustering*
James Gung and Jugal Kalita

11:20–11:40 *Comparing HMMs and Bayesian Networks for Surface Realisation*
Nina Dethlefs and Heriberto Cuayahuitl

11:40–12:00 *On The Feasibility of Open Domain Referring Expression Generation Using Large Scale Folksonomies*
Fabián Pacheco, Pablo Duboue and Martín Domínguez

Wednesday, June 6, 2012 (continued)

(12:00-1:00) Lunch

(1:00-2:00) Business Meeting

Session Wed-7E: (2:10-3:40) Sentiment and Social Media

2:10–2:40 *Structured Event Retrieval over Microblog Archives*
Donald Metzler, Congxing Cai and Eduard Hovy

2:40–3:10 *Learning from Bullying Traces in Social Media*
Jun-Ming Xu, Kwang-Sung Jun, Xiaojin Zhu and Amy Bellmore

3:10–3:40 *Grammatical structures for word-level sentiment detection*
Asad Sayeed, Jordan Boyd-Graber, Bryan Rusk and Amy Weinberg

Session Wed-7C: (2:10-3:40) Machine Learning II

2:10–2:40 *Graph-Based Lexicon Expansion with Sparsity-Inducing Penalties*
Dipanjan Das and Noah A. Smith

2:40–3:10 *Unified Expectation Maximization*
Rajhans Samdani, Ming-Wei Chang and Dan Roth

3:10–3:40 *Low-Dimensional Discriminative Reranking*
Jagadeesh Jagarlamudi and Hal Daume III

Wednesday, June 6, 2012 (continued)

Session Wed-7W: (2:10-3:40) Discourse, Dialog, and Pragmatics II

2:10–2:40 *Autonomous Self-Assessment of Autocorrections: Exploring Text Message Dialogues*
Tyler Baldwin and Joyce Chai

2:40–3:10 *Translation-Based Projection for Multilingual Coreference Resolution*
Altaf Rahman and Vincent Ng

3:10–3:40 *Exploring Semi-Supervised Coreference Resolution of Medical Concepts using Semantic and Temporal Features*
Preethi Raghavan, Eric Fosler-Lussier and Albert Lai

(3:40-4:10) Coffee Break

Session Wed-8E: (4:10-5:10) Summarization

4:10–4:40 *Mind the Gap: Learning to Choose Gaps for Question Generation*
Lee Becker, Sumit Basu and Lucy Vanderwende

4:40–5:10 *Unsupervised Concept-to-text Generation with Hypergraphs*
Ioannis Konstas and Mirella Lapata

Session Wed-8C: (4:10-5:10) Semantics II

4:10–4:40 *Detecting Visual Text*
Jesse Dodge, Amit Goyal, Xufeng Han, Alyssa Mensch, Margaret Mitchell, Karl Stratos, Kota Yamaguchi, Yejin Choi, Hal Daume III, Alex Berg and Tamara Berg

4:40–5:10 *Unsupervised Translation Sense Clustering*
Mohit Bansal, John DeNero and Dekang Lin

Session Wed-8W: (4:10-5:10) Document Categorization and Topic Modeling

4:10–4:40 *Shared Components Topic Models*
Matthew R. Gormley, Mark Dredze, Benjamin Van Durme and Jason Eisner

4:40–5:10 *Textual Predictors of Bill Survival in Congressional Committees*
Tae Yano, Noah A. Smith and John D. Wilkerson