

EMNLP 2013

**2013 Conference on Empirical Methods in
Natural Language Processing**

Proceedings of the Conference

18-21 October 2013
Grand Hyatt Seattle
Seattle, Washington, USA

We would like to thank our sponsors:

Platinum:



Gold:



ALLEN INSTITUTE
for ARTIFICIAL INTELLIGENCE

Microsoft[®]



NUANCE

Bronze:



IBM WATSON[™]

Supporter:



JOHNS HOPKINS
UNIVERSITY
human language technology
center of excellence

©2013 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 978-1-937284-97-8

Preface

Welcome to the 2013 Conference on Empirical Methods in Natural Language Processing.

EMNLP has grown to be one of the largest and most competitive conferences in computational linguistics. Organized by ACL SIGDAT (the Association for Computational Linguistics special interest group for linguistic data and corpus-based approaches to natural language processing), it features papers on all areas of interest to the SIGDAT community and aligned fields. It is being held this year as a standalone conference at the Grand Hyatt Seattle, in the heart of downtown Seattle, USA over the period October 18–21, 2013.

This year we introduced short papers to EMNLP for the first time, in an attempt to encourage submission of papers reporting smaller, more focused contributions and work in progress. We also put a lot of time and energy into “closing the loop” in the author response phase, in getting reviewers to explicitly acknowledge author responses and update their reviews where appropriate. We sincerely hope that this contributed towards further improvement in the quality of reviews and the decision-making process.

We received a record number of 772 valid submissions (not including co-submitted papers that were withdrawn from the conference), made up of 539 long papers and 233 short papers. These papers were reviewed across a total of 15 areas, of which Machine Translation (98 submissions), Semantics (87 submissions) and NLP-related Machine Learning (76) were the largest. The submissions were managed by 30 area chairs (two per area) and evaluated by a combined programme committee of 505 reviewers.

28% of the long paper submissions and 24% of the short paper submissions were accepted for publication at the conference. Five long papers were shortlisted for the best paper award, based on input from the reviewers and area chairs, and have been scheduled for presentation in a plenary session at the end of the conference, culminating in the presentation of the best paper award.

We would like to acknowledge all the hard work of the submitting authors, without whom there would, of course, be no conference. To the authors of accepted papers, we offer congratulations; to the authors of rejected papers, we offer our sincere commiserations, and dearly hope that the hard work of the programme committee provided you with valuable feedback on your research. We are eternally indebted to our dedicated and hard-working area chairs, and to the reviewers for their attention to detail and engagement with the author response/discussion phase, which was tremendously helpful in gauging the relative merits of each paper and being able to send out the notifications on time.

We are very grateful to our two invited speakers: Fernando Pereira (Research Director at Google) who will draw on his considerable experience and wisdom in presenting “Meaning in the Wild”, focusing on machine understanding; and Andrew Ng (Co-CEO and Co-founder of Coursera) who will discuss the challenges and opportunities associated with the delivery of Massively Open Online Courses (MOOCs) in a talk titled “The Online Revolution: Education for Everyone”, from the perspective of a true world leader in MOOC provision and Machine Learning/NLP research.

We would also like to thank the inimitable Priscilla Rasmussen who single-handedly looked after the local organisation of EMNLP 2013. We also wish to acknowledge the considerable efforts of Steven Berthard who put this volume together with peerless efficiency, Francesco Figari who took excellent

care of the conference website, and Rich Gerber from Softconf.com, who responded to any questions regarding START — the submission management system used for EMNLP 2013 — instantaneously and uncomplainingly, and helped us manage the large number of submissions smoothly. Additionally, we would like to thank Eugene Charniak, Mark Johnson and Noah Smith for serving on the best paper award committee, and providing characteristically probing and insightful critiques of the best paper award nominees.

Special thanks go to David Yarowsky, the general chair for the conference, who has provided us with much valuable advice, encouragement and assistance over the past six months. We would also like to thank the members of the SIGDAT board who advised us on various matters, and our predecessors James Henderson and Marius Pasca for nudging us in the right direction on a number of occasions.

On behalf of all attendees at the conference, we would also like to acknowledge the generosity of our sponsors/supports: Amazon, Google, the Allen Institute for Artificial Intelligence, Inome, IBM Research, Microsoft Research, Nuance and John Hopkins University.

It has been an honour to serve as Programme Chairs of EMNLP 2013. We sincerely hope that you — in equal measure — enjoy and are intellectually-stimulated by the conference, and have a pleasant stay in beautiful Seattle.

Timothy Baldwin and Anna Korhonen
EMNLP 2013 Programme Chairs

Organizers

General Chair:

David Yarowsky (John Hopkins University, USA)

Programme Chairs:

Timothy Baldwin (University of Melbourne, Australia)

Anna Korhonen (University of Cambridge, UK)

Workshops Chair:

Karen Livescu (Toyota Technological Institute at Chicago, USA)

Publication Chair:

Steven Bethard (University of Alabama at Birmingham, USA)

Area Chairs:

Phonology, Morphology, Tagging, Chunking and Segmentation:

Kemal Oflazer (Carnegie Mellon University Qatar)

Anna Feldman (Montclair State University, USA)

Syntax and Parsing:

Jennifer Foster (Dublin City University, Ireland)

Yoav Goldberg (Bar Ilan University, Israel)

Semantics:

Mark Stevenson (University of Sheffield, UK)

Luke Zettlemoyer (University of Washington, USA)

Discourse, Dialogue, and Pragmatics:

Carolyn Rose (Carnegie Mellon University, USA)

Matt Purver (Queen Mary, University of London, UK)

Language resources:

Emily M. Bender (University of Washington, USA)

Aline Villavicencio (Federal University of Rio Grande do Sul, Brazil)

Summarization and Generation:

Dragomir Radev (University of Michigan, USA)
Yang Liu (University of Texas at Dallas, USA)

NLP-related Machine Learning: theory, methods and algorithms:

Amir Globerson (The Hebrew University of Jerusalem, Israel)
Antal van den Bosch (Radboud University Nijmegen, Netherlands)

Machine Translation:

Taro Watanabe (NICT, Japan)
Kevin Knight (Information Sciences Institute, USA)

Information Retrieval and Question Answering:

Bernardo Magnini (Fondazione Bruno Kessler, Italy)
Soumen Chakrabarti (Indian Institute of Technology, India)

Information Extraction:

Mausam (Indian Institute of Technology, India)
Heng Ji (City University of New York, USA)

Spoken Language Processing:

Haizhou Li (Institute for Infocomm Research, Singapore)
Amanda Stent (AT&T Labs, USA)

Text Mining and Natural Language Processing Applications:

Hang Li (Huawei Noah's Ark Lab, Hong Kong)
Kevin Cohen (University of Colorado School of Medicine, USA)

Sentiment Analysis and Opinion Mining:

Janyce Weibe (University of Pittsburgh, USA)
Bing Liu (University of Illinois at Chicago, USA)

NLP for the Web and Social Media:

Miles Osborne (University of Edinburgh, UK)
Chin-Yew Lin (Microsoft Research Asia, China)

Computational Models of Human Language Acquisition and Processing:

Alessandro Lenci (University of Pisa, Italy)
Afra Alishahi (Tilburg University, Netherlands)

Table of Contents

<i>Event-Based Time Label Propagation for Automatic Dating of News Articles</i> Tao Ge, Baobao Chang, Sujian Li and Zhifang Sui	1
<i>Exploiting Discourse Analysis for Article-Wide Temporal Classification</i> Jun-Ping Ng, Min-Yen Kan, Ziheng Lin, Wei Feng, Bin Chen, Jian Su and Chew Lim Tan	12
<i>Combining Generative and Discriminative Model Scores for Distant Supervision</i> Benjamin Roth and Dietrich Klakow	24
<i>Exploring the Utility of Joint Morphological and Syntactic Learning from Child-directed Speech</i> Stella Frank, Frank Keller and Sharon Goldwater	30
<i>A Joint Learning Model of Word Segmentation, Lexical Acquisition, and Phonetic Variability</i> Micha Elsner, Sharon Goldwater, Naomi Feldman and Frank Wood	42
<i>Animacy Detection with Voting Models</i> Joshua Moore, Christopher J.C. Burges, Erin Renshaw and Wen-tau Yih	55
<i>A Log-Linear Model for Unsupervised Text Normalization</i> Yi Yang and Jacob Eisenstein	61
<i>Paraphrasing 4 Microblog Normalization</i> Wang Ling, Chris Dyer, Alan W Black and Isabel Trancoso	73
<i>Question Difficulty Estimation in Community Question Answering Services</i> Jing Liu, Quan Wang, Chin-Yew Lin and Hsiao-Wuen Hon	85
<i>Measuring Ideological Proportions in Political Speeches</i> Yanchuan Sim, Brice D. L. Acree, Justin H. Gross and Noah A. Smith	91
<i>Learning to Freestyle: Hip Hop Challenge-Response Induction via Transduction Rule Segmentation</i> Dekai Wu, Kartek Addanki, Markus Saers and Meriem Beloucif	102
<i>Modeling Scientific Impact with Topical Influence Regression</i> James Foulds and Padhraic Smyth	113
<i>Joint Parsing and Disfluency Detection in Linear Time</i> Mohammad Sadegh Rasooli and Joel Tetreault	124
<i>Modeling and Learning Semantic Co-Compositionality through Prototype Projections and Neural Networks</i> Masashi Tsubaki, Kevin Duh, Masashi Shimbo and Yuji Matsumoto	130
<i>Studying the Recursive Behaviour of Adjectival Modification with Compositional Distributional Semantics</i> Eva Maria Vecchi, Roberto Zamparelli and Marco Baroni	141

<i>Learning Latent Word Representations for Domain Adaptation using Supervised Word Clustering</i> Min Xiao, Feipeng Zhao and Yuhong Guo	152
<i>Appropriately Incorporating Statistical Significance in PMI</i> Om P. Damani and Shweta Ghonge	163
<i>Growing Multi-Domain Glossaries from a Few Seeds using Probabilistic Topic Models</i> Stefano Faralli and Roberto Navigli	170
<i>Joint Learning of Phonetic Units and Word Pronunciations for ASR</i> Chia-ying Lee, Yu Zhang and James Glass	182
<i>MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text</i> Matthew Richardson, Christopher J.C. Burges and Erin Renshaw	193
<i>Noise-Aware Character Alignment for Bootstrapping Statistical Machine Transliteration from Bilingual Corpora</i> Katsuhito Sudoh, Shinsuke Mori and Masaaki Nagata	204
<i>Optimal Beam Search for Machine Translation</i> Alexander Rush, Yin-Wen Chang and Michael Collins	210
<i>An Efficient Language Model Using Double-Array Structures</i> Makoto Yasuhara, Toru Tanaka, Jun-ya Norimatsu and Mikio Yamamoto	222
<i>Structured Penalties for Log-Linear Language Models</i> Anil Kumar Nelakanti, Cedric Archambeau, Julien Mairal, Francis Bach and Guillaume Bouchard	233
<i>Interactive Machine Translation using Hierarchical Translation Models</i> Jesús González-Rubio, Daniel Ortíz-Martínez, José-Miguel Benedí and Francisco Casacuberta	244
<i>Max-Margin Synchronous Grammar Induction for Machine Translation</i> Xinyan Xiao and Deyi Xiong	255
<i>Error-Driven Analysis of Challenges in Coreference Resolution</i> Jonathan K. Kummerfeld and Dan Klein	265
<i>Exploiting Zero Pronouns to Improve Chinese Coreference Resolution</i> Fang Kong and Hwee Tou Ng	278
<i>Joint Coreference Resolution and Named-Entity Linking with Multi-Pass Sieves</i> Hannaneh Hajishirzi, Leila Zilles, Daniel S. Weld and Luke Zettlemoyer	289
<i>Interpreting Anaphoric Shell Nouns using Antecedents of Cataphoric Shell Nouns as Training Data</i> Varada Kolhatkar, Heike Zinsmeister and Graeme Hirst	300
<i>Exploring Representations from Unlabeled Data with Co-training for Chinese Word Segmentation</i> Longkai Zhang, Houfeng Wang, Xu Sun and Mairgup Mansur	311

<i>Efficient Higher-Order CRFs for Morphological Tagging</i> Thomas Mueller, Helmut Schmid and Hinrich Schütze	322
<i>The Effects of Syntactic Features in Automatic Prediction of Morphology</i> Wolfgang Seeker and Jonas Kuhn	333
<i>Adaptor Grammars for Learning Non-Concatenative Morphology</i> Jan A. Botha and Phil Blunsom	345
<i>Grounding Strategic Conversation: Using Negotiation Dialogues to Predict Trades in a Win-Lose Game</i> Anais Cadilhac, Nicholas Asher, Farah Benamara and Alex Lascarides	357
<i>Unsupervised Induction of Contingent Event Pairs from Film Scenes</i> Zhichao Hu, Elahe Rahimtoroghi, Larissa Munishkina, Reid Swanson and Marilyn A. Walker	369
<i>Latent Anaphora Resolution for Cross-Lingual Pronoun Prediction</i> Christian Hardmeier, Jörg Tiedemann and Joakim Nivre	380
<i>Towards Situated Dialogue: Revisiting Referring Expression Generation</i> Rui Fang, Changsong Liu, Lanbo She and Joyce Y. Chai	392
<i>Open-Domain Fine-Grained Class Extraction from Web Search Queries</i> Marius Pasca	403
<i>Unsupervised Relation Extraction with General Domain Knowledge</i> Oier Lopez de Lacalle and Mirella Lapata	415
<i>Efficient Collective Entity Linking with Stacking</i> Zhengyan He, Shujie Liu, Yang Song, Mu Li, Ming Zhou and Houfeng Wang	426
<i>Joint Bootstrapping of Corpus Annotations and Entity Types</i> Hrushikesh Mohapatra, Siddhanth Jain and Soumen Chakrabarti	436
<i>Effectiveness and Efficiency of Open Relation Extraction</i> Filipe Mesquita, Jordan Schmeidek and Denilson Barbosa	447
<i>Automatic Feature Engineering for Answer Selection and Extraction</i> Aliaksei Severyn and Alessandro Moschitti	458
<i>Improving Web Search Ranking by Incorporating Structured Annotation of Queries</i> Xiao Ding, Zhicheng Dou, Bing Qin, Ting Liu and Ji-rong Wen	468
<i>Building Specialized Bilingual Lexicons Using Large Scale Background Knowledge</i> Dhouha Bouamor, Adrian Popescu, Nasredine Semmar and Pierre Zweigenbaum	479
<i>Document Summarization via Guided Sentence Compression</i> Chen Li, Fei Liu, Fuliang Weng and Yang Liu	490
<i>Anchor Graph: Global Reordering Contexts for Statistical Machine Translation</i> Hendra Setiawan, Bowen Zhou and Bing Xiang	501

<i>Source-Side Classifier Preordering for Machine Translation</i> Uri Lerner and Slav Petrov	513
<i>Improving Pivot-Based Statistical Machine Translation Using Random Walk</i> Xiaoning Zhu, Zhongjun He, Hua Wu, Haifeng Wang, Conghui Zhu and Tiejun Zhao	524
<i>Improving Alignment of System Combination by Using Multi-objective Optimization</i> Tian Xia, Zongcheng Ji, Shaodan Zhai, Yidong Chen, Qun Liu and Shaojun Wang	535
<i>Flexible and Efficient Hypergraph Interactions for Joint Hierarchical and Forest-to-String Decoding</i> Martin Cmejrek, Haitao Mi and Bowen Zhou	545
<i>Factored Soft Source Syntactic Constraints for Hierarchical Machine Translation</i> Zhongqiang Huang, Jacob Devlin and Rabih Zbib	556
<i>Recursive Autoencoders for ITG-Based Translation</i> Peng Li, Yang Liu and Maosong Sun	567
<i>Automatically Classifying Edit Categories in Wikipedia Revisions</i> Johannes Daxenberger and Iryna Gurevych	578
<i>Semi-Markov Phrase-Based Monolingual Alignment</i> Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch and Peter Clark	590
<i>A Constrained Latent Variable Model for Coreference Resolution</i> Kai-Wei Chang, Rajhans Samdani and Dan Roth	601
<i>Centering Similarity Measures to Reduce Hubs</i> Ikumi Suzuki, Kazuo Hara, Masashi Shimbo, Marco Saerens and Kenji Fukumizu	613
<i>Unsupervised Spectral Learning of WCFG as Low-rank Matrix Completion</i> Raphaël Bailly, Xavier Carreras, Franco M. Luque and Ariadna Quattoni	624
<i>Identifying Phrasal Verbs Using Many Bilingual Corpora</i> Karl Pichotta and John DeNero	636
<i>Deep Learning for Chinese Word Segmentation and POS Tagging</i> Xiaoqing Zheng, Hanyang Chen and Tianyu Xu	647
<i>Joint Chinese Word Segmentation and POS Tagging on Heterogeneous Annotated Corpora with Multiple Task Learning</i> Xipeng Qiu, Jiayi Zhao and Xuanjing Huang	658
<i>The Topology of Semantic Knowledge</i> Jimmy Dubuisson, Jean-Pierre Eckmann, Christian Scheible and Hinrich Schütze	669
<i>Unsupervised Induction of Cross-Lingual Semantic Relations</i> Mike Lewis and Mark Steedman	681

<i>Two-Stage Method for Large-Scale Acquisition of Contradiction Pattern Pairs using Entailment</i> Julien Kloetzer, Stijn De Saeger, Kentaro Torisawa, Chikara Hashimoto, Jong-Hoon Oh, Motoki Sano and Kiyonori Ohtake	693
<i>Sarcasm as Contrast between a Positive Sentiment and Negative Situation</i> Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert and Ruihong Huang	704
<i>Collective Personal Profile Summarization with Social Networks</i> Zhongqing Wang, Shoushan LI, Fang Kong and Guodong Zhou	715
<i>Optimized Event Storyline Generation based on Mixture-Event-Aspect Model</i> Lifu Huang and Lian'en Huang	726
<i>Automatically Determining a Proper Length for Multi-Document Summarization: A Bayesian Non-parametric Approach</i> Tengfei Ma and Hiroshi Nakagawa	736
<i>A Discourse-Driven Content Model for Summarising Scientific Articles Evaluated in a Complex Question Answering Task</i> Maria Liakata, Simon Dobnik, Shyamasree Saha, Colin Batchelor and Dietrich Rebbholz-Schuhmann	747
<i>Optimal Incremental Parsing via Best-First Dynamic Programming</i> Kai Zhao, James Cross and Liang Huang	758
<i>Exploiting Language Models for Visual Recognition</i> Dieu-Thu Le, Jasper Uijlings and Raffaella Bernardi	769
<i>Mining Scientific Terms and their Definitions: A Study of the ACL Anthology</i> Yiping Jin, Min-Yen Kan, Jun-Ping Ng and Xiangnan He	780
<i>Joint Learning and Inference for Grammatical Error Correction</i> Alla Rozovskaya and Dan Roth	791
<i>With Blinkers on: Robust Prediction of Eye Movements across Readers</i> Franz Matthies and Anders Søgaard	803
<i>Using Paraphrases and Lexical Semantics to Improve the Accuracy and the Robustness of Supervised Models in Situated Dialogue Systems</i> Claire Gardent and Lina M. Rojas Barahona	808
<i>Cascading Collective Classification for Bridging Anaphora Recognition using a Rich Linguistic Feature Set</i> Yufang Hou, Katja Markert and Michael Strube	814
<i>A Synchronous Context Free Grammar for Time Normalization</i> Steven Bethard	821

<i>Rule-Based Information Extraction is Dead! Long Live Rule-Based Information Extraction Systems!</i> Laura Chiticariu, Yunyao Li and Frederick R. Reiss	827
<i>Improving Learning and Inference in a Large Knowledge-Base using Latent Syntactic Cues</i> Matt Gardner, Partha Pratim Talukdar, Bryan Kisiel and Tom Mitchell	833
<i>What is Hidden among Translation Rules</i> Libin Shen and Bowen Zhou	839
<i>Converting Continuous-Space Language Models into N-Gram Language Models for Statistical Machine Translation</i> Rui Wang, Masao Utiyama, Isao Goto, Eiichro Sumita, Hai Zhao and Bao-Liang Lu	845
<i>A Corpus Level MIRA Tuning Strategy for Machine Translation</i> Ming Tan, Tian Xia, Shaojun Wang and Bowen Zhou	851
<i>Word Level Language Identification in Online Multilingual Communication</i> Dong Nguyen and A. Seza Dogruoz	857
<i>Microblog Entity Linking by Leveraging Extra Posts</i> Yuhang Guo, Bing Qin, Ting Liu and Sheng Li	863
<i>Automatic Domain Partitioning for Multi-Domain Learning</i> Di Wang, Chenyan Xiong and William Yang Wang	869
<i>Decipherment with a Million Random Restarts</i> Taylor Berg-Kirkpatrick and Dan Klein	874
<i>Russian Stress Prediction using Maximum Entropy Ranking</i> Keith Hall and Richard Sproat	879
<i>Scaling to Large³ Data: An Efficient and Effective Method to Compute Distributional Thesauri</i> Martin Riedl and Chris Biemann	884
<i>Discriminative Improvements to Distributional Sentence Similarity</i> Yangfeng Ji and Jacob Eisenstein	891
<i>Is Twitter A Better Corpus for Measuring Sentiment Similarity?</i> Shi Feng, Le Zhang, Binyang Li, Daling Wang, Ge Yu and Kam-Fai Wong	897
<i>Implicit Feature Detection via a Constrained Topic Model and SVM</i> Wei Wang, Hua Xu and Xiaoqiu Huang	903
<i>Online Learning for Inexact Hypergraph Search</i> Hao Zhang, Liang Huang, Kai Zhao and Ryan McDonald	908
<i>Predicting the Presence of Discourse Connectives</i> Gary Patterson and Andrew Kehler	914

<i>Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions</i>	
Masatsugu Hangyo, Daisuke Kawahara and Sadao Kurohashi	924
<i>A Dataset for Research on Short-Text Conversations</i>	
Hao Wang, Zhengdong Lu, Hang Li and Enhong Chen	935
<i>Discourse Level Explanatory Relation Extraction from Product Reviews Using First-Order Logic</i>	
Qi Zhang, Jin Qian, Huan Chen, Jihua Kang and Xuanjing Huang	946
<i>Building Event Threads out of Multiple News Articles</i>	
Xavier Tannier and Véronique Moriceau	958
<i>Tree Kernel-based Negation and Speculation Scope Detection with Structured Syntactic Parse Features</i>	
Bowei Zou, Guodong Zhou and Qiaoming Zhu	968
<i>A temporal model of text periodicities using Gaussian Processes</i>	
Daniel Preoțiu-Pietro and Trevor Cohn	977
<i>Automatically Detecting and Attributing Indirect Quotations</i>	
Silvia Pareti, Tim O’Keefe, Ioannis Konstas, James R. Curran and Irena Koprinska	989
<i>Identifying Web Search Query Reformulation using Concept based Matching</i>	
Ahmed Hassan	1000
<i>The Answer is at your Fingertips: Improving Passage Retrieval for Web Question Answering with Search Behavior Data</i>	
Mikhail Ageev, Dmitry Lagun and Eugene Agichtein	1011
<i>Assembling the Kazakh Language Corpus</i>	
Olzhas Makhambetov, Aibek Makazhanov, Zhandos Yessenbayev, Bakhyt Matkarimov, Islam Sabyrgaliyev and Anuar Sharafudinov	1022
<i>Automatic Extraction of Morphological Lexicons from Morphologically Annotated Corpora</i>	
Ramy Eskander, Nizar Habash and Owen Rambow	1032
<i>Joint Language and Translation Modeling with Recurrent Neural Networks</i>	
Michael Auli, Michel Galley, Chris Quirk and Geoffrey Zweig	1044
<i>Multi-Domain Adaptation for SMT Using Multi-Task Learning</i>	
Lei Cui, Xilun Chen, Dongdong Zhang, Shujie Liu, Mu Li and Ming Zhou	1055
<i>Translation with Source Constituency and Dependency Trees</i>	
Fandong Meng, Jun Xie, Linfeng Song, Yajuan Lü and Qun Liu	1066
<i>Monolingual Marginal Matching for Translation Model Adaptation</i>	
Ann Irvine, Chris Quirk and Hal Daumé III	1077
<i>Efficient Left-to-Right Hierarchical Phrase-Based Translation with Improved Reordering</i>	
Maryam Siahbani, Baskaran Sankaran and Anoop Sarkar	1089

<i>A Systematic Exploration of Diversity in Machine Translation</i>	
Kevin Gimpel, Dhruv Batra, Chris Dyer and Gregory Shakhnarovich	1100
<i>Max-Violation Perceptron and Forced Decoding for Scalable MT Training</i>	
Heng Yu, Liang Huang, Haitao Mi and Kai Zhao	1112
<i>Identifying Multiple Userids of the Same Author</i>	
Tieyun Qian and Bing Liu	1124
<i>Gender Inference of Twitter Users in Non-English Contexts</i>	
Morgane Ciot, Morgan Sonderegger and Derek Ruths	1136
<i>A Multimodal LDA Model integrating Textual, Cognitive and Visual Modalities</i>	
Stephen Roller and Sabine Schulte im Walde	1146
<i>Combining PCFG-LA Models with Dual Decomposition: A Case Study with Function Labels and Binarization</i>	
Joseph Le Roux, Antoine Rozenknop and Jennifer Foster	1158
<i>Feature Noising for Log-Linear Structured Prediction</i>	
Sida Wang, Mengqiu Wang, Stefan Wager, Percy Liang and Christopher D. Manning	1170
<i>Improvements to the Bayesian Topic N-Gram Models</i>	
Hiroshi Noji, Daichi Mochihashi and Yusuke Miyao	1180
<i>An Empirical Study Of Semi-Supervised Chinese Word Segmentation Using Co-Training</i>	
Fan Yang and Paul Vozila	1191
<i>Ubertagging: Joint Segmentation and Supertagging for English</i>	
Rebecca Dridan	1201
<i>Automatic Knowledge Acquisition for Case Alternation between the Passive and Active Voices in Japanese</i>	
Ryohei Sasano, Daisuke Kawahara, Sadao Kurohashi and Manabu Okumura	1213
<i>Exploiting Multiple Sources for Open-Domain Hypernym Discovery</i>	
Ruiji Fu, Bing Qin and Ting Liu	1224
<i>A Semantically Enhanced Approach to Determine Textual Similarity</i>	
Eduardo Blanco and Dan Moldovan	1235
<i>Understanding and Quantifying Creativity in Lexical Composition</i>	
Polina Kuznetsova, Jianfu Chen and Yejin Choi	1246
<i>Sentiment Analysis: How to Derive Prior Polarities from SentiWordNet</i>	
Marco Guerini, Lorenzo Gatti and Marco Turchi	1259
<i>Simulating Early-Termination Search for Verbose Spoken Queries</i>	
Jerome White, Douglas W. Oard, Nitendra Rajput and Marion Zalk	1270

<i>Summarizing Complex Events: a Cross-Modal Solution of Storylines Extraction and Reconstruction</i> Shize Xu, Shanshan Wang and Yan Zhang	1281
<i>Image Description using Visual Dependency Representations</i> Desmond Elliott and Frank Keller	1292
<i>Semi-Supervised Feature Transformation for Dependency Parsing</i> Wenliang Chen, Min Zhang and Yue Zhang	1303
<i>Leveraging Lexical Cohesion and Disruption for Topic Segmentation</i> Anca-Roxana Simon, Guillaume Gravier and Pascale Sébillot	1314
<i>This Text Has the Scent of Starbucks: A Laplacian Structured Sparsity Model for Computational Branding Analytics</i> William Yang Wang, Edward Lin and John Kominek	1325
<i>Mining New Business Opportunities: Identifying Trend related Products by Leveraging Commercial Intents from Microblogs</i> Jinpeng Wang, Wayne Xin Zhao, Haitian Wei, Hongfei Yan and Xiaoming Li	1337
<i>Using Topic Modeling to Improve Prediction of Neuroticism and Depression in College Students</i> Philip Resnik, Anderson Garron and Rebecca Resnik	1348
<i>Predicting the Resolution of Referring Expressions from User Behavior</i> Nikos Engonopoulos, Martin Villalba, Ivan Titov and Alexander Koller	1354
<i>Chinese Zero Pronoun Resolution: Some Recent Advances</i> Chen Chen and Vincent Ng	1360
<i>Connecting Language and Knowledge Bases with Embedding Models for Relation Extraction</i> Jason Weston, Antoine Bordes, Oksana Yakhnenko and Nicolas Usunier	1366
<i>Simple Customization of Recursive Neural Networks for Semantic Relation Classification</i> Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka and Takashi Chikayama	1372
<i>Improving Statistical Machine Translation with Word Class Models</i> Joern Wuebker, Stephan Peitz, Felix Rietig and Hermann Ney	1377
<i>Shift-Reduce Word Reordering for Machine Translation</i> Katsuhiko Hayashi, Katsuhito Sudoh, Hajime Tsukada, Jun Suzuki and Masaaki Nagata ...	1382
<i>Decoding with Large-Scale Neural Language Models Improves Translation</i> Ashish Vaswani, Yinggong Zhao, Victoria Fossum and David Chiang	1387
<i>Bilingual Word Embeddings for Phrase-Based Machine Translation</i> Will Y. Zou, Richard Socher, Daniel Cer and Christopher D. Manning	1393
<i>Application of Localized Similarity for Web Documents</i> Peter Reberšek and Mateja Verlic	1399

<i>Dependency Language Models for Sentence Completion</i>	
Joseph Gubbins and Andreas Vlachos	1405
<i>A Walk-Based Semantically Enriched Tree Kernel Over Distributed Word Representations</i>	
Shashank Srivastava, Dirk Hovy and Eduard Hovy	1411
<i>Automatic Idiom Identification in Wiktionary</i>	
Grace Muzny and Luke Zettlemoyer	1417
<i>Elephant: Sequence Labeling for Word and Sentence Segmentation</i>	
Kilian Evang, Valerio Basile, Grzegorz Chrupala and Johan Bos	1422
<i>Detecting Compositionality of Multi-Word Expressions using Nearest Neighbours in Vector Space Models</i>	
Douwe Kiela and Stephen Clark	1427
<i>Naive Bayes Word Sense Induction</i>	
Do Kook Choe and Eugene Charniak	1433
<i>The VerbCorner Project: Toward an Empirically-Based Semantic Decomposition of Verbs</i>	
Joshua K. Hartshorne, Claire Bonial and Martha Palmer	1438
<i>Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews</i>	
Jun Seok Kang, Polina Kuznetsova, Michael Luca and Yejin Choi	1443
<i>Automatically Identifying Pseudepigraphic Texts</i>	
Moshe Koppel and Shachar Seidman	1449
<i>Dynamic Feature Selection for Dependency Parsing</i>	
He He, Hal Daumé III and Jason Eisner	1455
<i>Semi-Supervised Representation Learning for Cross-Lingual Text Classification</i>	
Min Xiao and Yuhong Guo	1465
<i>Using Crowdsourcing to get Representations based on Regular Expressions</i>	
Anders Søgaard, Hector Martinez, Jakob Elming and Anders Johannsen	1476
<i>Overcoming the Lack of Parallel Data in Sentence Compression</i>	
Katja Filippova and Yasemin Altun	1481
<i>Fast Joint Compression and Summarization via Graph Cuts</i>	
Xian Qian and Yang Liu	1492
<i>Inducing Document Plans for Concept-to-Text Generation</i>	
Ioannis Konstas and Mirella Lapata	1503
<i>Single-Document Summarization as a Tree Knapsack Problem</i>	
Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda and Masaaki Nagata ..	1515

<i>A Hierarchical Entity-Based Approach to Structuralize User Generated Content in Social Media: A Case of Yahoo! Answers</i>	
Baichuan Li, Jing Liu, Chin-Yew Lin, Irwin King and Michael R. Lyu	1521
<i>Semantic Parsing on Freebase from Question-Answer Pairs</i>	
Jonathan Berant, Andrew Chou, Roy Frostig and Percy Liang	1533
<i>Scaling Semantic Parsers with On-the-Fly Ontology Matching</i>	
Tom Kwiatkowski, Eunsol Choi, Yoav Artzi and Luke Zettlemoyer	1545
<i>Classifying Message Board Posts with an Extracted Lexicon of Patient Attributes</i>	
Ruihong Huang and Ellen Riloff	1557
<i>Lexical Chain Based Cohesion Models for Document-Level Statistical Machine Translation</i>	
Deyi Xiong, Yang Ding, Min Zhang and Chew Lim Tan	1563
<i>A Convex Alternative to IBM Model 2</i>	
Andrei Simion, Michael Collins and Cliff Stein	1574
<i>Pair Language Models for Deriving Alternative Pronunciations and Spellings from Pronunciation Dictionaries</i>	
Russell Beckley and Brian Roark	1584
<i>Prior Disambiguation of Word Tensors for Constructing Sentence Vectors</i>	
Dimitri Kartsaklis and Mehrnoosh Sadrzadeh	1590
<i>Multi-Relational Latent Semantic Analysis</i>	
Kai-Wei Chang, Wen-tau Yih and Christopher Meek	1602
<i>A Study on Bootstrapping Bilingual Vector Spaces from Non-Parallel Data (and Nothing Else)</i>	
Ivan Vulić and Marie-Francine Moens	1613
<i>Deriving Adjectival Scales from Continuous Space Word Representations</i>	
Joo-Kyung Kim and Marie-Catherine de Marneffe	1625
<i>Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank</i>	
Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng and Christopher Potts	1631
<i>Open Domain Targeted Sentiment</i>	
Margaret Mitchell, Jacqui Aguilar, Theresa Wilson and Benjamin Van Durme	1643
<i>Exploiting Domain Knowledge in Aspect Extraction</i>	
Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos and Riddhiman Ghosh	1655
<i>Dependency-Based Decipherment for Resource-Limited Machine Translation</i>	
Qing Dou and Kevin Knight	1668

<i>Translating into Morphologically Rich Languages with Synthetic Phrases</i> Victor Chahuneau, Eva Schlinger, Noah A. Smith and Chris Dyer	1677
<i>Boosting Cross-Language Retrieval by Learning Bilingual Phrase Associations from Relevance Rankings</i> Artem Sokokov, Laura Jehl, Felix Hieber and Stefan Riezler	1688
<i>Recurrent Continuous Translation Models</i> Nal Kalchbrenner and Phil Blunsom	1700
<i>Learning Biological Processes with Global Constraints</i> Aju Thalappillil Scaria, Jonathan Berant, Mengqiu Wang, Peter Clark, Justin Lewis, Brittany Harding and Christopher D. Manning	1710
<i>Generating Coherent Event Schemas at Scale</i> Niranjana Balasubramanian, Stephen Soderland, Mausam and Oren Etzioni	1721
<i>Orthonormal Explicit Topic Analysis for Cross-Lingual Document Matching</i> John Philip McCrae, Philipp Cimiano and Roman Klinger	1732
<i>Automated Essay Scoring by Maximizing Human-Machine Agreement</i> Hongbo Chen and Ben He	1741
<i>Success with Style: Using Writing Style to Predict the Success of Novels</i> Vikas Ganjigunte Ashok, Song Feng and Yejin Choi	1753
<i>A Generative Joint, Additive, Sequential Model of Topics and Speech Acts in Patient-Doctor Communication</i> Byron C. Wallace, Thomas A Trikalinos, M. Barton Laws, Ira B. Wilson and Eugene Charniak	1765
<i>Harvesting Parallel News Streams to Generate Paraphrases of Event Relations</i> Congle Zhang and Daniel S. Weld	1776
<i>Relational Inference for Wikification</i> Xiao Cheng and Dan Roth	1787
<i>Event Schema Induction with a Probabilistic Entity-Driven Model</i> Nathanael Chambers	1797
<i>Using Soft Constraints in Joint Inference for Clinical Concept Recognition</i> Prateek Jindal and Dan Roth	1808
<i>Exploring Demographic Language Variations to Improve Multilingual Sentiment Analysis in Social Media</i> Svitlana Volkova, Theresa Wilson and David Yarowsky	1815
<i>Opinion Mining in Newspaper Articles by Entropy-Based Word Connections</i> Thomas Scholz and Stefan Conrad	1828

<i>Collective Opinion Target Extraction in Chinese Microblogs</i> Xinjie Zhou, Xiaojun Wan and Jianguo Xiao	1840
<i>Detecting Promotional Content in Wikipedia</i> Shruti Bhosale, Heath Vinicombe and Raymond Mooney	1851
<i>Learning Topics and Positions from Debatepedia</i> Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang and Noah A. Smith	1858
<i>A Unified Model for Topics, Events and Users on Twitter</i> Qiming Diao and Jing Jiang	1869
<i>Authorship Attribution of Micro-Messages</i> Roy Schwartz, Oren Tsur, Ari Rappoport and Moshe Koppel	1880
<i>Detection of Product Comparisons - How Far Does an Out-of-the-Box Semantic Role Labeling System Take You?</i> Wiltrud Kessler and Jonas Kuhn	1892
<i>A Multi-Teraflop Constituency Parser using GPUs</i> John Canny, David Hall and Dan Klein	1898
<i>Fish Transporters and Miracle Homes: How Compositional Distributional Semantics can Help NP Parsing</i> Angeliki Lazaridou, Eva Maria Vecchi and Marco Baroni	1908
<i>Learning Distributions over Logical Forms for Referring Expression Generation</i> Nicholas FitzGerald, Yoav Artzi and Luke Zettlemoyer	1914
<i>Learning to Rank Lexical Substitutions</i> György Szarvas, Róbert Busa-Fekete and Eyke Hüllermeier	1926
<i>Identifying Manipulated Offerings on Review Portals</i> Jiwei Li, Myle Ott and Claire Cardie	1933
<i>Well-Argued Recommendation: Adaptive Models Based on Words in Recommender Systems</i> Julien Gaillard, Marc El-Beze, Eitan Altman and Emmanuel Ethis	1943
<i>Regularized Minimum Error Rate Training</i> Michel Galley, Chris Quirk, Colin Cherry and Kristina Toutanova	1948
<i>Of Words, Eyes and Brains: Correlating Image-Based Distributional Semantic Models with Neural Representations of Concepts</i> Andrew J. Anderson, Elia Bruni, Ulisse Bordignon, Massimo Poesio and Marco Baroni	1960
<i>Easy Victories and Uphill Battles in Coreference Resolution</i> Greg Durrett and Dan Klein	1971

Breaking Out of Local Optima with Count Transforms and Model Recombination: A Study in Grammar Induction

Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky 1983

Cross-Lingual Discriminative Learning of Sequence Models with Posterior Regularization

Kuzman Ganchev and Dipanjan Das 1996

Conference Program

Saturday, October 19, 2013

(7:30-9:00) Breakfast

(9:00-9:15) Opening Remarks

(9:20-10:30) Information Extraction I

9:20–9:45 *Event-Based Time Label Propagation for Automatic Dating of News Articles*
Tao Ge, Baobao Chang, Sujian Li and Zhifang Sui

9:45–10:10 *Exploiting Discourse Analysis for Article-Wide Temporal Classification*
Jun-Ping Ng, Min-Yen Kan, Ziheng Lin, Wei Feng, Bin Chen, Jian Su and Chew Lim Tan

10:10–10:30 *Combining Generative and Discriminative Model Scores for Distant Supervision*
Benjamin Roth and Dietrich Klakow

(9:20-10:30) Language Acquisition and Processing

9:20–9:45 *Exploring the Utility of Joint Morphological and Syntactic Learning from Child-directed Speech*
Stella Frank, Frank Keller and Sharon Goldwater

9:45–10:10 *A Joint Learning Model of Word Segmentation, Lexical Acquisition, and Phonetic Variability*
Micha Elsner, Sharon Goldwater, Naomi Feldman and Frank Wood

10:10–10:30 *Animacy Detection with Voting Models*
Joshua Moore, Christopher J.C. Burges, Erin Renshaw and Wen-tau Yih

Saturday, October 19, 2013 (continued)

(9:20-10:30) NLP for Social Media I

- 9:20–9:45 *A Log-Linear Model for Unsupervised Text Normalization*
Yi Yang and Jacob Eisenstein
- 9:45–10:10 *Paraphrasing 4 Microblog Normalization*
Wang Ling, Chris Dyer, Alan W Black and Isabel Trancoso
- 10:10–10:30 *Question Difficulty Estimation in Community Question Answering Services*
Jing Liu, Quan Wang, Chin-Yew Lin and Hsiao-Wuen Hon
- (10:30-
11:00) Break

(11:00-12:35) NLP Applications I

- 11:00–11:25 *Measuring Ideological Proportions in Political Speeches*
Yanchuan Sim, Brice D. L. Acree, Justin H. Gross and Noah A. Smith
- 11:25–11:50 *Learning to Freestyle: Hip Hop Challenge-Response Induction via Transduction Rule Segmentation*
Dekai Wu, Karteek Addanki, Markus Saers and Meriem Beloucif
- 11:50–12:15 *Modeling Scientific Impact with Topical Influence Regression*
James Foulds and Padhraic Smyth
- 12:15–12:35 *Joint Parsing and Disfluency Detection in Linear Time*
Mohammad Sadegh Rasooli and Joel Tetreault

Saturday, October 19, 2013 (continued)

(11:00-12:35) Semantics I

- 11:00–11:25 *Modeling and Learning Semantic Co-Compositionality through Prototype Projections and Neural Networks*
Masashi Tsubaki, Kevin Duh, Masashi Shimbo and Yuji Matsumoto
- 11:25–11:50 *Studying the Recursive Behaviour of Adjectival Modification with Compositional Distributional Semantics*
Eva Maria Vecchi, Roberto Zamparelli and Marco Baroni
- 11:50–12:15 *Learning Latent Word Representations for Domain Adaptation using Supervised Word Clustering*
Min Xiao, Feipeng Zhao and Yuhong Guo
- 12:15–12:35 *Appropriately Incorporating Statistical Significance in PMI*
Om P. Damani and Shweta Ghonge

(11:00-12:35) Language Resources

- 11:00–11:25 *Growing Multi-Domain Glossaries from a Few Seeds using Probabilistic Topic Models*
Stefano Faralli and Roberto Navigli
- 11:25–11:50 *Joint Learning of Phonetic Units and Word Pronunciations for ASR*
Chia-ying Lee, Yu Zhang and James Glass
- 11:50–12:15 *MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text*
Matthew Richardson, Christopher J.C. Burges and Erin Renshaw
- 12:15–12:35 *Noise-Aware Character Alignment for Bootstrapping Statistical Machine Transliteration from Bilingual Corpora*
Katsuhito Sudoh, Shinsuke Mori and Masaaki Nagata
- (12:35–2:00) Lunch
- (2:00–3:15) First invited talk: Andrew Ng
- (3:15–3:45) Break

Saturday, October 19, 2013 (continued)

(3:45-5:50) Machine Translation I

- 3:45–4:10 *Optimal Beam Search for Machine Translation*
Alexander Rush, Yin-Wen Chang and Michael Collins
- 4:10–4:35 *An Efficient Language Model Using Double-Array Structures*
Makoto Yasuhara, Toru Tanaka, Jun-ya Norimatsu and Mikio Yamamoto
- 4:35–5:00 *Structured Penalties for Log-Linear Language Models*
Anil Kumar Nelakanti, Cedric Archambeau, Julien Mairal, Francis Bach and Guillaume Bouchard
- 5:00–5:25 *Interactive Machine Translation using Hierarchical Translation Models*
Jesús González-Rubio, Daniel Ortíz-Martínez, José-Miguel Benedí and Francisco Casacuberta
- 5:25–5:50 *Max-Margin Synchronous Grammar Induction for Machine Translation*
Xinyan Xiao and Deyi Xiong

(3:45-5:50) Dialogue and Discourse

- 3:45–4:10 *Error-Driven Analysis of Challenges in Coreference Resolution*
Jonathan K. Kummerfeld and Dan Klein
- 4:10–4:35 *Exploiting Zero Pronouns to Improve Chinese Coreference Resolution*
Fang Kong and Hwee Tou Ng
- 4:35–5:00 *Joint Coreference Resolution and Named-Entity Linking with Multi-Pass Sieves*
Hannaneh Hajishirzi, Leila Zilles, Daniel S. Weld and Luke Zettlemoyer
- 5:00–5:25 *Interpreting Anaphoric Shell Nouns using Antecedents of Cataphoric Shell Nouns as Training Data*
Varada Kolhatkar, Heike Zinsmeister and Graeme Hirst

Saturday, October 19, 2013 (continued)

(3:45-5:50) Morphology and Phonology

- 3:45–4:10 *Exploring Representations from Unlabeled Data with Co-training for Chinese Word Segmentation*
Longkai Zhang, Houfeng Wang, Xu Sun and Mairgup Mansur
- 4:10–4:35 *Efficient Higher-Order CRFs for Morphological Tagging*
Thomas Mueller, Helmut Schmid and Hinrich Schütze
- 4:35–5:00 *The Effects of Syntactic Features in Automatic Prediction of Morphology*
Wolfgang Seeker and Jonas Kuhn
- 5:00–5:25 *Adaptor Grammars for Learning Non-Concatenative Morphology*
Jan A. Botha and Phil Blunsom

(6:00-7:30) Poster Session A

Grounding Strategic Conversation: Using Negotiation Dialogues to Predict Trades in a Win-Lose Game

Anais Cadilhac, Nicholas Asher, Farah Benamara and Alex Lascarides

Unsupervised Induction of Contingent Event Pairs from Film Scenes

Zhichao Hu, Elahe Rahimtoroghi, Larissa Munishkina, Reid Swanson and Marilyn A. Walker

Latent Anaphora Resolution for Cross-Lingual Pronoun Prediction

Christian Hardmeier, Jörg Tiedemann and Joakim Nivre

Towards Situated Dialogue: Revisiting Referring Expression Generation

Rui Fang, Changsong Liu, Lanbo She and Joyce Y. Chai

Open-Domain Fine-Grained Class Extraction from Web Search Queries

Marius Pasca

Unsupervised Relation Extraction with General Domain Knowledge

Oier Lopez de Lacalle and Mirella Lapata

Efficient Collective Entity Linking with Stacking

Zhengyan He, Shujie Liu, Yang Song, Mu Li, Ming Zhou and Houfeng Wang

Saturday, October 19, 2013 (continued)

Joint Bootstrapping of Corpus Annotations and Entity Types

Hrushikesh Mohapatra, Siddhanth Jain and Soumen Chakrabarti

Effectiveness and Efficiency of Open Relation Extraction

Filipe Mesquita, Jordan Schmidek and Denilson Barbosa

Automatic Feature Engineering for Answer Selection and Extraction

Aliaksei Severyn and Alessandro Moschitti

Improving Web Search Ranking by Incorporating Structured Annotation of Queries

Xiao Ding, Zhicheng Dou, Bing Qin, Ting Liu and Ji-rong Wen

Building Specialized Bilingual Lexicons Using Large Scale Background Knowledge

Dhouha Bouamor, Adrian Popescu, Nasredine Semmar and Pierre Zweigenbaum

Document Summarization via Guided Sentence Compression

Chen Li, Fei Liu, Fuliang Weng and Yang Liu

Anchor Graph: Global Reordering Contexts for Statistical Machine Translation

Hendra Setiawan, Bowen Zhou and Bing Xiang

Source-Side Classifier Preordering for Machine Translation

Uri Lerner and Slav Petrov

Improving Pivot-Based Statistical Machine Translation Using Random Walk

Xiaoning Zhu, Zhongjun He, Hua Wu, Haifeng Wang, Conghui Zhu and Tiejun Zhao

Improving Alignment of System Combination by Using Multi-objective Optimization

Tian Xia, Zongcheng Ji, Shaodan Zhai, Yidong Chen, Qun Liu and Shaojun Wang

Flexible and Efficient Hypergraph Interactions for Joint Hierarchical and Forest-to-String Decoding

Martin Cmejrek, Haitao Mi and Bowen Zhou

Factored Soft Source Syntactic Constraints for Hierarchical Machine Translation

Zhongqiang Huang, Jacob Devlin and Rabih Zbib

Saturday, October 19, 2013 (continued)

Recursive Autoencoders for ITG-Based Translation

Peng Li, Yang Liu and Maosong Sun

Automatically Classifying Edit Categories in Wikipedia Revisions

Johannes Daxenberger and Iryna Gurevych

Semi-Markov Phrase-Based Monolingual Alignment

Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch and Peter Clark

A Constrained Latent Variable Model for Coreference Resolution

Kai-Wei Chang, Rajhans Samdani and Dan Roth

Centering Similarity Measures to Reduce Hubs

Ikumi Suzuki, Kazuo Hara, Masashi Shimbo, Marco Saerens and Kenji Fukumizu

Unsupervised Spectral Learning of WCFG as Low-rank Matrix Completion

Raphaël Bailly, Xavier Carreras, Franco M. Luque and Ariadna Quattoni

Identifying Phrasal Verbs Using Many Bilingual Corpora

Karl Pichotta and John DeNero

Deep Learning for Chinese Word Segmentation and POS Tagging

Xiaoqing Zheng, Hanyang Chen and Tianyu Xu

Joint Chinese Word Segmentation and POS Tagging on Heterogeneous Annotated Corpora with Multiple Task Learning

Xipeng Qiu, Jiayi Zhao and Xuanjing Huang

The Topology of Semantic Knowledge

Jimmy Dubuisson, Jean-Pierre Eckmann, Christian Scheible and Hinrich Schütze

Unsupervised Induction of Cross-Lingual Semantic Relations

Mike Lewis and Mark Steedman

Two-Stage Method for Large-Scale Acquisition of Contradiction Pattern Pairs using Entailment

Julien Kloetzer, Stijn De Saeger, Kentaro Torisawa, Chikara Hashimoto, Jong-Hoon Oh, Motoki Sano and Kiyonori Ohtake

Saturday, October 19, 2013 (continued)

Sarcasm as Contrast between a Positive Sentiment and Negative Situation

Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert and Ruihong Huang

Collective Personal Profile Summarization with Social Networks

Zhongqing Wang, Shoushan LI, Fang Kong and Guodong Zhou

Optimized Event Storyline Generation based on Mixture-Event-Aspect Model

Lifu Huang and Lian'en Huang

Automatically Determining a Proper Length for Multi-Document Summarization: A Bayesian Nonparametric Approach

Tengfei Ma and Hiroshi Nakagawa

A Discourse-Driven Content Model for Summarising Scientific Articles Evaluated in a Complex Question Answering Task

Maria Liakata, Simon Dobnik, Shyamasree Saha, Colin Batchelor and Dietrich Rebholz-Schuhmann

Optimal Incremental Parsing via Best-First Dynamic Programming

Kai Zhao, James Cross and Liang Huang

Exploiting Language Models for Visual Recognition

Dieu-Thu Le, Jasper Uijlings and Raffaella Bernardi

Mining Scientific Terms and their Definitions: A Study of the ACL Anthology

Yiping Jin, Min-Yen Kan, Jun-Ping Ng and Xiangnan He

Joint Learning and Inference for Grammatical Error Correction

Alla Rozovskaya and Dan Roth

With Blinkers on: Robust Prediction of Eye Movements across Readers

Franz Matthies and Anders Søgaard

Using Paraphrases and Lexical Semantics to Improve the Accuracy and the Robustness of Supervised Models in Situated Dialogue Systems

Claire Gardent and Lina M. Rojas Barahona

Cascading Collective Classification for Bridging Anaphora Recognition using a Rich Linguistic Feature Set

Yufang Hou, Katja Markert and Michael Strube

Saturday, October 19, 2013 (continued)

A Synchronous Context Free Grammar for Time Normalization

Steven Bethard

Rule-Based Information Extraction is Dead! Long Live Rule-Based Information Extraction Systems!

Laura Chiticariu, Yunyao Li and Frederick R. Reiss

Improving Learning and Inference in a Large Knowledge-Base using Latent Syntactic Cues

Matt Gardner, Partha Pratim Talukdar, Bryan Kisiel and Tom Mitchell

What is Hidden among Translation Rules

Libin Shen and Bowen Zhou

Converting Continuous-Space Language Models into N-Gram Language Models for Statistical Machine Translation

Rui Wang, Masao Utiyama, Isao Goto, Eiichiro Sumita, Hai Zhao and Bao-Liang Lu

A Corpus Level MIRA Tuning Strategy for Machine Translation

Ming Tan, Tian Xia, Shaojun Wang and Bowen Zhou

Word Level Language Identification in Online Multilingual Communication

Dong Nguyen and A. Seza Dogruoz

Microblog Entity Linking by Leveraging Extra Posts

Yuhang Guo, Bing Qin, Ting Liu and Sheng Li

Automatic Domain Partitioning for Multi-Domain Learning

Di Wang, Chenyan Xiong and William Yang Wang

Decipherment with a Million Random Restarts

Taylor Berg-Kirkpatrick and Dan Klein

Russian Stress Prediction using Maximum Entropy Ranking

Keith Hall and Richard Sproat

Scaling to Large³ Data: An Efficient and Effective Method to Compute Distributional Thesauri

Martin Riedl and Chris Biemann

Saturday, October 19, 2013 (continued)

Discriminative Improvements to Distributional Sentence Similarity

Yangfeng Ji and Jacob Eisenstein

Is Twitter A Better Corpus for Measuring Sentiment Similarity?

Shi Feng, Le Zhang, Binyang Li, Daling Wang, Ge Yu and Kam-Fai Wong

Implicit Feature Detection via a Constrained Topic Model and SVM

Wei Wang, Hua Xu and Xiaoqiu Huang

Online Learning for Inexact Hypergraph Search

Hao Zhang, Liang Huang, Kai Zhao and Ryan McDonald

(7:30-9:00) Poster Session B

Predicting the Presence of Discourse Connectives

Gary Patterson and Andrew Kehler

Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions

Masatsugu Hangyo, Daisuke Kawahara and Sadao Kurohashi

A Dataset for Research on Short-Text Conversations

Hao Wang, Zhengdong Lu, Hang Li and Enhong Chen

Discourse Level Explanatory Relation Extraction from Product Reviews Using First-Order Logic

Qi Zhang, Jin Qian, Huan Chen, Jihua Kang and Xuanjing Huang

Building Event Threads out of Multiple News Articles

Xavier Tannier and Véronique Moriceau

Tree Kernel-based Negation and Speculation Scope Detection with Structured Syntactic Parse Features

Bowei Zou, Guodong Zhou and Qiaoming Zhu

A temporal model of text periodicities using Gaussian Processes

Daniel Preoțiuc-Pietro and Trevor Cohn

Automatically Detecting and Attributing Indirect Quotations

Silvia Pareti, Tim O'Keefe, Ioannis Konstas, James R. Curran and Irena Koprinska

Saturday, October 19, 2013 (continued)

Identifying Web Search Query Reformulation using Concept based Matching

Ahmed Hassan

The Answer is at your Fingertips: Improving Passage Retrieval for Web Question Answering with Search Behavior Data

Mikhail Ageev, Dmitry Lagun and Eugene Agichtein

Assembling the Kazakh Language Corpus

Olzhas Makhambetov, Aibek Makazhanov, Zhandos Yessenbayev, Bakhyt Matkarimov, Islam Sabyrgaliyev and Anuar Sharafudinov

Automatic Extraction of Morphological Lexicons from Morphologically Annotated Corpora

Ramy Eskander, Nizar Habash and Owen Rambow

Joint Language and Translation Modeling with Recurrent Neural Networks

Michael Auli, Michel Galley, Chris Quirk and Geoffrey Zweig

Multi-Domain Adaptation for SMT Using Multi-Task Learning

Lei Cui, Xilun Chen, Dongdong Zhang, Shujie Liu, Mu Li and Ming Zhou

Translation with Source Constituency and Dependency Trees

Fandong Meng, Jun Xie, Linfeng Song, Yajuan Lü and Qun Liu

Monolingual Marginal Matching for Translation Model Adaptation

Ann Irvine, Chris Quirk and Hal Daumé III

Efficient Left-to-Right Hierarchical Phrase-Based Translation with Improved Reordering

Maryam Siahbani, Baskaran Sankaran and Anoop Sarkar

A Systematic Exploration of Diversity in Machine Translation

Kevin Gimpel, Dhruv Batra, Chris Dyer and Gregory Shakhnarovich

Max-Violation Perceptron and Forced Decoding for Scalable MT Training

Heng Yu, Liang Huang, Haitao Mi and Kai Zhao

Identifying Multiple Userids of the Same Author

Tieyun Qian and Bing Liu

Saturday, October 19, 2013 (continued)

Gender Inference of Twitter Users in Non-English Contexts

Morgane Ciot, Morgan Sonderegger and Derek Ruths

A Multimodal LDA Model integrating Textual, Cognitive and Visual Modalities

Stephen Roller and Sabine Schulte im Walde

Combining PCFG-LA Models with Dual Decomposition: A Case Study with Function Labels and Binarization

Joseph Le Roux, Antoine Rozenknop and Jennifer Foster

Feature Noising for Log-Linear Structured Prediction

Sida Wang, Mengqiu Wang, Stefan Wager, Percy Liang and Christopher D. Manning

Improvements to the Bayesian Topic N-Gram Models

Hiroshi Noji, Daichi Mochihashi and Yusuke Miyao

An Empirical Study Of Semi-Supervised Chinese Word Segmentation Using Co-Training

Fan Yang and Paul Vozila

Ubertagging: Joint Segmentation and Supertagging for English

Rebecca Dridan

Automatic Knowledge Acquisition for Case Alternation between the Passive and Active Voices in Japanese

Ryohei Sasano, Daisuke Kawahara, Sadao Kurohashi and Manabu Okumura

Exploiting Multiple Sources for Open-Domain Hypernym Discovery

Ruiji Fu, Bing Qin and Ting Liu

A Semantically Enhanced Approach to Determine Textual Similarity

Eduardo Blanco and Dan Moldovan

Understanding and Quantifying Creativity in Lexical Composition

Polina Kuznetsova, Jianfu Chen and Yejin Choi

Sentiment Analysis: How to Derive Prior Polarities from SentiWordNet

Marco Guerini, Lorenzo Gatti and Marco Turchi

Saturday, October 19, 2013 (continued)

Simulating Early-Termination Search for Verbose Spoken Queries

Jerome White, Douglas W. Oard, Nitendra Rajput and Marion Zalk

Summarizing Complex Events: a Cross-Modal Solution of Storylines Extraction and Re-construction

Shize Xu, Shanshan Wang and Yan Zhang

Image Description using Visual Dependency Representations

Desmond Elliott and Frank Keller

Semi-Supervised Feature Transformation for Dependency Parsing

Wenliang Chen, Min Zhang and Yue Zhang

Leveraging Lexical Cohesion and Disruption for Topic Segmentation

Anca-Roxana Simon, Guillaume Gravier and Pascale Sébillot

This Text Has the Scent of Starbucks: A Laplacian Structured Sparsity Model for Computational Branding Analytics

William Yang Wang, Edward Lin and John Kominek

Mining New Business Opportunities: Identifying Trend related Products by Leveraging Commercial Intents from Microblogs

Jinpeng Wang, Wayne Xin Zhao, Haitian Wei, Hongfei Yan and Xiaoming Li

Using Topic Modeling to Improve Prediction of Neuroticism and Depression in College Students

Philip Resnik, Anderson Garron and Rebecca Resnik

Predicting the Resolution of Referring Expressions from User Behavior

Nikos Engonopoulos, Martin Villalba, Ivan Titov and Alexander Koller

Chinese Zero Pronoun Resolution: Some Recent Advances

Chen Chen and Vincent Ng

Connecting Language and Knowledge Bases with Embedding Models for Relation Extraction

Jason Weston, Antoine Bordes, Oksana Yakhnenko and Nicolas Usunier

Simple Customization of Recursive Neural Networks for Semantic Relation Classification

Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka and Takashi Chikayama

Saturday, October 19, 2013 (continued)

Improving Statistical Machine Translation with Word Class Models

Joern Wuebker, Stephan Peitz, Felix Rietig and Hermann Ney

Shift-Reduce Word Reordering for Machine Translation

Katsuhiko Hayashi, Katsuhito Sudoh, Hajime Tsukada, Jun Suzuki and Masaaki Nagata

Decoding with Large-Scale Neural Language Models Improves Translation

Ashish Vaswani, Yinggong Zhao, Victoria Fossum and David Chiang

Bilingual Word Embeddings for Phrase-Based Machine Translation

Will Y. Zou, Richard Socher, Daniel Cer and Christopher D. Manning

Application of Localized Similarity for Web Documents

Peter Reberšek and Mateja Verlic

Dependency Language Models for Sentence Completion

Joseph Gubbins and Andreas Vlachos

A Walk-Based Semantically Enriched Tree Kernel Over Distributed Word Representations

Shashank Srivastava, Dirk Hovy and Eduard Hovy

Automatic Idiom Identification in Wiktionary

Grace Muzny and Luke Zettlemoyer

Elephant: Sequence Labeling for Word and Sentence Segmentation

Kilian Evang, Valerio Basile, Grzegorz Chrupała and Johan Bos

Detecting Compositionality of Multi-Word Expressions using Nearest Neighbours in Vector Space Models

Douwe Kiela and Stephen Clark

Naive Bayes Word Sense Induction

Do Kook Choe and Eugene Charniak

The VerbCorner Project: Toward an Empirically-Based Semantic Decomposition of Verbs

Joshua K. Hartshorne, Claire Bonial and Martha Palmer

Saturday, October 19, 2013 (continued)

Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews

Jun Seok Kang, Polina Kuznetsova, Michael Luca and Yejin Choi

Automatically Identifying Pseudepigraphic Texts

Moshe Koppel and Shachar Seidman

Sunday, October 20, 2013

(8:00-9:00) Breakfast

(9:15-10:30) Second invited talk: Fernando Pereira

(10:30-11:00) Break

(11:00-12:35) Machine Learning for NLP

11:25–11:50 *Dynamic Feature Selection for Dependency Parsing*

He He, Hal Daumé III and Jason Eisner

11:50–12:15 *Semi-Supervised Representation Learning for Cross-Lingual Text Classification*

Min Xiao and Yuhong Guo

12:15–12:35 *Using Crowdsourcing to get Representations based on Regular Expressions*

Anders Søgaard, Hector Martinez, Jakob Elming and Anders Johannsen

(11:00-12:35) Summarization and Generation

11:00–11:25 *Overcoming the Lack of Parallel Data in Sentence Compression*

Katja Filippova and Yasemin Altun

11:25–11:50 *Fast Joint Compression and Summarization via Graph Cuts*

Xian Qian and Yang Liu

11:50–12:15 *Inducing Document Plans for Concept-to-Text Generation*

Ioannis Konstas and Mirella Lapata

Sunday, October 20, 2013 (continued)

12:15–12:35 *Single-Document Summarization as a Tree Knapsack Problem*
Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda and Masaaki Nagata

(11:00-12:35) Information Extraction and Social Media Analysis

11:00–11:25 *A Hierarchical Entity-Based Approach to Structuralize User Generated Content in Social Media: A Case of Yahoo! Answers*
Baichuan Li, Jing Liu, Chin-Yew Lin, Irwin King and Michael R. Lyu

11:25–11:50 *Semantic Parsing on Freebase from Question-Answer Pairs*
Jonathan Berant, Andrew Chou, Roy Frostig and Percy Liang

11:50–12:15 *Scaling Semantic Parsers with On-the-Fly Ontology Matching*
Tom Kwiatkowski, Eunsol Choi, Yoav Artzi and Luke Zettlemoyer

12:15–12:35 *Classifying Message Board Posts with an Extracted Lexicon of Patient Attributes*
Ruihong Huang and Ellen Riloff

(12:35-2:00) Lunch

(2:00-3:35) Machine Translation II

2:00–2:25 *Lexical Chain Based Cohesion Models for Document-Level Statistical Machine Translation*
Deyi Xiong, Yang Ding, Min Zhang and Chew Lim Tan

2:50–3:15 *A Convex Alternative to IBM Model 2*
Andrei Simion, Michael Collins and Cliff Stein

3:15–3:35 *Pair Language Models for Deriving Alternative Pronunciations and Spellings from Pronunciation Dictionaries*
Russell Beckley and Brian Roark

Sunday, October 20, 2013 (continued)

(2:00-3:35) Semantics II

- 2:00–2:25 *Prior Disambiguation of Word Tensors for Constructing Sentence Vectors*
Dimitri Kartsaklis and Mehrnoosh Sadrzadeh
- 2:25–2:50 *Multi-Relational Latent Semantic Analysis*
Kai-Wei Chang, Wen-tau Yih and Christopher Meek
- 2:50–3:15 *A Study on Bootstrapping Bilingual Vector Spaces from Non-Parallel Data (and Nothing Else)*
Ivan Vulić and Marie-Francine Moens
- 3:15–3:35 *Deriving Adjectival Scales from Continuous Space Word Representations*
Joo-Kyung Kim and Marie-Catherine de Marneffe

(2:00-3:35) Opinion Mining and Sentiment Analysis I

- 2:00–2:25 *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*
Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng and Christopher Potts
- 2:25–2:50 *Open Domain Targeted Sentiment*
Margaret Mitchell, Jacqui Aguilar, Theresa Wilson and Benjamin Van Durme
- 2:50–3:15 *Exploiting Domain Knowledge in Aspect Extraction*
Zhiyuan Chen, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos and Riddhiman Ghosh
- (3:35-4:05) Break

Sunday, October 20, 2013 (continued)

(4:05-5:55) Machine Translation III

- 4:05–4:30 *Dependency-Based Decipherment for Resource-Limited Machine Translation*
Qing Dou and Kevin Knight
- 4:30–4:55 *Translating into Morphologically Rich Languages with Synthetic Phrases*
Victor Chahuneau, Eva Schlinger, Noah A. Smith and Chris Dyer
- 4:55–5:20 *Boosting Cross-Language Retrieval by Learning Bilingual Phrase Associations from Relevance Rankings*
Artem Sokolov, Laura Jehl, Felix Hieber and Stefan Riezler
- 5:20–5:55 *Recurrent Continuous Translation Models*
Nal Kalchbrenner and Phil Blunsom

(4:05-5:55) Information Extraction II

- 4:05–4:30 *Learning Biological Processes with Global Constraints*
Aju Thalappillil Scaria, Jonathan Berant, Mengqiu Wang, Peter Clark, Justin Lewis, Brit-tany Harding and Christopher D. Manning
- 4:30–4:55 *Generating Coherent Event Schemas at Scale*
Niranjan Balasubramanian, Stephen Soderland, Mausam and Oren Etzioni
- 5:20–5:55 *Orthonormal Explicit Topic Analysis for Cross-Lingual Document Matching*
John Philip McCrae, Philipp Cimiano and Roman Klinger

(4:05-5:55) NLP Applications II

- 4:05–4:30 *Automated Essay Scoring by Maximizing Human-Machine Agreement*
Hongbo Chen and Ben He
- 4:55–5:20 *Success with Style: Using Writing Style to Predict the Success of Novels*
Vikas Ganjigunte Ashok, Song Feng and Yejin Choi
- 5:20–5:55 *A Generative Joint, Additive, Sequential Model of Topics and Speech Acts in Patient-Doctor Communication*
Byron C. Wallace, Thomas A Trikalinos, M. Barton Laws, Ira B. Wilson and Eugene Charniak

Monday, October 21, 2013

(8:00-9:00) Breakfast

(9:00-10:35) Information Extraction III

9:00–9:25 *Harvesting Parallel News Streams to Generate Paraphrases of Event Relations*
Congle Zhang and Daniel S. Weld

9:25–9:50 *Relational Inference for Wikification*
Xiao Cheng and Dan Roth

9:50–10:15 *Event Schema Induction with a Probabilistic Entity-Driven Model*
Nathanael Chambers

10:15–10:35 *Using Soft Constraints in Joint Inference for Clinical Concept Recognition*
Prateek Jindal and Dan Roth

(9:00-10:35) Opinion Mining and Sentiment Analysis II

9:00–9:25 *Exploring Demographic Language Variations to Improve Multilingual Sentiment Analysis in Social Media*
Svitlana Volkova, Theresa Wilson and David Yarowsky

9:25–9:50 *Opinion Mining in Newspaper Articles by Entropy-Based Word Connections*
Thomas Scholz and Stefan Conrad

9:50–10:15 *Collective Opinion Target Extraction in Chinese Microblogs*
Xinjie Zhou, Xiaojun Wan and Jianguo Xiao

10:15–10:35 *Detecting Promotional Content in Wikipedia*
Shruti Bhosale, Heath Vinicombe and Raymond Mooney

Monday, October 21, 2013 (continued)

(9:00-10:35) NLP for Social Media II

- 9:00–9:25 *Learning Topics and Positions from Debatepedia*
Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang and Noah A. Smith
- 9:25–9:50 *A Unified Model for Topics, Events and Users on Twitter*
Qiming Diao and Jing Jiang
- 9:50–10:15 *Authorship Attribution of Micro-Messages*
Roy Schwartz, Oren Tsur, Ari Rappoport and Moshe Koppel
- 10:15–10:35 *Detection of Product Comparisons - How Far Does an Out-of-the-Box Semantic Role Labeling System Take You?*
Wiltrud Kessler and Jonas Kuhn

(10:35–11:00) Break

(11:00-11:45) Parsing

- 11:00–11:25 *A Multi-Teraflop Constituency Parser using GPUs*
John Canny, David Hall and Dan Klein
- 11:25–11:45 *Fish Transporters and Miracle Homes: How Compositional Distributional Semantics can Help NP Parsing*
Angeliki Lazaridou, Eva Maria Vecchi and Marco Baroni

(11:00-11:45) Semantics III

- 11:00–11:25 *Learning Distributions over Logical Forms for Referring Expression Generation*
Nicholas FitzGerald, Yoav Artzi and Luke Zettlemoyer
- 11:25–11:45 *Learning to Rank Lexical Substitutions*
György Szarvas, Róbert Busa-Fekete and Eyke Hüllermeier

Monday, October 21, 2013 (continued)

(11:00-11:45) NLP Applications III

11:00–11:25 *Identifying Manipulated Offerings on Review Portals*

Jiwei Li, Myle Ott and Claire Cardie

11:25–11:45 *Well-Argued Recommendation: Adaptive Models Based on Words in Recommender Systems*

Julien Gaillard, Marc El-Beze, Eitan Altman and Emmanuel Ethis

(11:45–12:15) SIGDAT business meeting

(12:15–1:45) Lunch

(1:45–3:15) Plenary session I

1:45–2:15 *Regularized Minimum Error Rate Training*

Michel Galley, Chris Quirk, Colin Cherry and Kristina Toutanova

2:15–2:45 *Of Words, Eyes and Brains: Correlating Image-Based Distributional Semantic Models with Neural Representations of Concepts*

Andrew J. Anderson, Elia Bruni, Ulisse Bordignon, Massimo Poesio and Marco Baroni

2:45–3:15 *Easy Victories and Uphill Battles in Coreference Resolution*

Greg Durrett and Dan Klein

(3:15–3:45) Break

(3:45–4:45) Plenary Session II

3:45–4:15 *Breaking Out of Local Optima with Count Transforms and Model Recombination: A Study in Grammar Induction*

Valentin I. Spitzkovsky, Hiyan Alshawi and Daniel Jurafsky

4:15–4:45 *Cross-Lingual Discriminative Learning of Sequence Models with Posterior Regularization*

Kuzman Ganchev and Dipanjan Das

(4:45–5:15) Closing session

