EMNLP 2008

2008 Conference on Empirical Methods in Natural Language Processing

Proceedings of the Conference

25–27 October 2008
Honolulu, Hawaii, USA

A meeting of SIGDAT, a Special Interest Group of the ACL

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

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

Preface

Welcome to the 2008 Conference on Empirical Methods in Natural Language Processing! The conference is organized under the auspices of SIGDAT, the ACL Special Interest Group for linguistic data and corpus-based approaches to natural language processing. It is co-located this year with AMTA 2008 and the International Workshop on Spoken Language Translation, in Honolulu, Hawaii.

EMNLP received 385 submissions. We were able to accept 116 papers in total (an acceptance rate of 30%). 81 of the papers (21%) were accepted for oral presentation, and 35 (9%) for poster presentation. Two poster papers were subsequently withdrawn after acceptance. The papers were selected by a program committee of 15 area chairs, from Asia, Europe, and North America, assisted by a panel of 339 reviewers. This year EMNLP introduced an author response period. Authors were able to read and respond to the reviews of their paper before the program committee made a final decision. They were asked to correct factual errors in the reviews and answer questions raised in the reviewer comments. The intention was to help produce more accurate reviews. In some cases, reviewers changed their scores in view of the authors' response and the area chairs read all responses carefully prior to making recommendations for acceptance.

First and foremost, we would like to thank the authors who submitted their work to EMNLP. The sheer number of submissions reflects how broad and active our field is. We are deeply indebted to the area chairs and the reviewers for their hard work. They enabled us to select an exciting program and to provide valuable feedback to the authors. We are grateful to our invited speakers Oren Etzioni, Tom Griffiths, and Fernando Pereira who graciously agreed to give talks at EMNLP. Additional thanks to the Publications Chair, Sebastian Padó, who put this volume together. Jason Eisner helped us immensely by compiling a web site on "How to Serve as Program Chair of a Conference" (`http://www.cs.jhu.edu/~jason/advice/how-to-chair-a-conference.html`). Special thanks to David Yarowsky and Ken Church of SIGDAT who provided much valuable advice and assistance over the past months. David also helped raise important financial support for the conference. We are most grateful to Priscilla Rasmussen who helped us with various logistic and organizational aspects of the conference. Rich Gerber and the START team responded to our questions quickly, and helped us manage the large number of submissions smoothly. Finally, thanks are due to our webmaster, Francesco Figari, who revamped our conference website on very short notice.

We hope you enjoy the conference!

Mirella Lapata and Hwee Tou Ng
EMNLP 2008 Program Co-Chairs

Program Co-Chairs:

Mirella Lapata, University of Edinburgh
Hwee Tou Ng, National University of Singapore

Area Chairs:

Eneko Agirre, University of the Basque Country
Srinivas Bangalore, AT&T Research
Noemie Elhadad, Columbia University
Radu Florian, IBM Research
Rebecca Hwa, University of Pittsburgh
Dan Klein, University of California at Berkeley
Hang Li, Microsoft Research
Mu Li, Microsoft Research
Jimmy Lin, University of Maryland
Chris Manning, Stanford University
Yuji Matsumoto, Nara Institute of Science and Technology
Mari Ostendorf, University of Washington
Bo Pang, Yahoo! Research
Chris Quirk, Microsoft Research
Ben Taskar, University of Pennsylvania

Local Arrangements Chair:

Priscilla Rasmussen

Publications Chair:

Sebastian Padó, Stanford University

Reviewers:

Meni Adler, Eugene Agichtein, Amr Ahmed, Yaser Al-Onaizan, Galen Andrew;

Peter Bailey, Timothy Baldwin, Carmen Banea, Roy Bar-Haim, Regina Barzilay, Roberto Basili, Cosmin Adrian Bejan, Anja Belz, Daniel Bikel, Misha Bilenko, Alexandra Birch, Alan Black, Elizabeth Boschee, Jordan Boyd-Graber, Eric Breck, Peter Bruza, Ivan Bulyko, Razvan Bunescu, Harry Bunt, John Burger, Bill Byrne, Donna Byron;

Chris Callison-Burch, Nicoletta Calzolari, Claire Cardie, Xavier Carreras, Vittorio Castelli, Neus Català, Dan Cer, Özgür Çetin, Nate Chambers, Pi-Chuan Chang, Eugene Charniak, Ciprian Chelba, Hsin-Hsi Chen, Colin Cherry, David Chiang, Yejin Choi, Jennifer Chu-Carroll, Tat-Seng Chua, Ken Church, Massimiliano Ciaramita, Alexander Clark, Stephen Clark, James Clarke, Michael Collins, Koby Crammer, Mathias Creutz, Dan Cristea, Silviu Cucerzan, Hang Cui;

Jan Daciuk, Walter Daelemans, Sajib Dasgupta, Hal Daume, Marie-Catherine de Marneffe, Maarten de Rijke, Steve DeNeefe, John DeNero, Mona Diab, Shilin Ding, Bill Dolan, Mark Dras, Mark Dredze, Kevin Duh, Chris Dyer, Myroslava Dzikovska;

Phil Edmonds, Jason Eisner, Michael Elhadad, Katrin Erk;

Hui Fang, Marcello Federico, Raquel Fernandez, Jenny Finkel, Alex Fraser, Marjorie Freedman, Dayne Freitag, Atsushi Fujii, Pascale Fung;

Ryan Gabbard, Robert Gaizauskas, Michel Galley, Michael Gamon, Kuzman Ganchev, Jianfeng Gao, Ruifang Ge, Darren Gergle, Ulrich Germann, Dan Gildea, Jonathan Ginzburg, Roxana Girju, Amir Globerson, Sharon Goldwater, Joao Graca, Ralph Grishman;

Nizar Habash, Aria Haghighi, Udo Hahn, Dilek Hakkani-Tur, Keith Hall, Jirka Hana, Sanda Harabagiu, Mary Harper, Mark Hasegawa-Johnson, Timothy Hazen, Xiaodong He, James Henderson, John Henderson, Ulf Hermjakob, Andrew Hickl, Ryuichiro Higashinaka, Dustin Hillard, Graeme Hirst, Julia Hockenmaier, Beth Ann Hockey, Véronique Hoste, Wu Hua, Jimmy Huang, Liang Huang;

Nancy Ide, Diana Inkpen, Hideki Isozaki, Abe Ittycheriah;

Heng Ji, Jing Jiang, Valentin Jijkoun, Howard Johnson, Mark Johnson, Rie Johnson, Kristiina Jokinen, Pamela Jordan, Joemon Jose, Aravind Joshi;

Michael Kaisser, Min-Yen Kan, Hiroshi Kanayama, Noriko Kando, Nikiforos Karamanis, Rohit Kate, Junichi Kazama, Frank Keller, Sanjeev Khudanpur, Katrin Kirchhoff, Kevin Knight, Philipp Koehn, Alexander Koller, Terry Koo, Moshe Koppel, Valia Kordoni, Anna Korhonen, Taku Kudo, Sandra Kuebler, Roland Kuhn, Alex Kulesza, Ravi Kumar, Shankar Kumar;

Simon Lacoste-Julien, Wai Lam, Philippe Langlais, Guy Lapalme, Guy Lebanon, Gary Geunbae Lee, Lillian Lee, Gina Levow, Roger Levy, Chi-Ho Li, Xiao Li, Zhifei Li, Percy Liang, Ee-Peng Lim, Kenneth Litkowski, Bing Liu, Yang Liu, Yang Liu, Karen Livescu, Jose Luis Vicedo, Xiaoqiang Luo, Yajuan Lv;

Bill MacCartney, Bernardo Magnini, Gideon Mann, Daniel Marcu, Katja Markert, David Martinez, Arne Mauser, Diana McCarthy, David McClosky, David McDonald, Ryan McDonald, Kathy McKeown, Michael McTear, Arul Menezes, Donald Metzler, Rada Mihalcea, Gilad Mishne, Teruko Mitamura, Diego Molla Aliod, Christof Monz, Robert Moore, Alessandro Moschitti;

Tetsuji Nakagawa, Satoshi Nakamura, Preslav Nakov, Vivi Nastase, Roberto Navigli, Ani Nenkova, Vincent Ng, Grace Ngai, Patrick Nguyen, Gabriel Nicolae;

Franz Och, Kemal Oflazer, Arzucan Özgür;

Sebastian Padó, Tim Paek, Martha Palmer, Patrick Pantel, Marius Pasca, Marco Pennacchiotti, Slav Petrov, Joseph Picone, Ana-Maria Popescu, Victor Poznanski, Sameer Pradhan, John Prager, Kathrin Probst, Vasin Punyakanok;

Tao Qin;

Dan Ramage, Owen Rambow, Deepak Ravichandran, Norbert Reithinger, Philip Resnik, German Rigau, Hae-Chang Rim, Fabio Rinaldi, Brian Roark, Patrick Ruch;

Yoshinori Sagisaka, Magnus Sahlgren, Tetsuya Sakai, David Schlangen, Helmut Schmid, Holger Schwenk, Frédérique Segond, Yohei Seki, Satoshi Sekine, Stephanie Seneff, Izhak Shafran, Libin Shen, Takahiro Shinozaki, Advaith Siddharthan, Khalil Sima'an, Manhung Siu, David Smith, Noah Smith, Rion Snow, Dawei Song, Mark Steedman, Amanda Stent, Mark Stevenson, Veselin Stoyanov, Michael Strube, Le Sun, Maosong Sun, Mihai Surdeanu, Charles Sutton, Stan Szpakowicz, Idan Szpektor;

Hiroya Takamura, Partha Pratim Talukdar, Marta Tatu, Simone Teufel, Christoph Tillmann, Ivan Titov, David Traum, Reut Tsarfaty, Huihsin Tseng, Junichi Tsujii, Dan Tufiş, Gökhan Tür, Joseph Turian, Evelyne Tzoukermann;

Kiyotaka Uchimoto, Takehito Utsuro;

Lucy Vanderwende, Stephan Vogel, Ellen Voorhees, Piek Vossen;

Stephen Wan, Qin Wang, Wei Wang, Wen Wang, Ye-Yi Wang, Taro Watanabe, Bonnie Webber, David Weir, Ralph Weischedel, Michael White, Richard Wicentowski, Dominic Widdows, Jason Williams, Theresa Wilson, Shuly Wintner, Yuk Wah Wong, Britta Wrede, Dekai Wu;

Fei Xia, Jinxi Xu, Jun Xu, Peng Xu, Guirong Xue;

Mei Yang, Qiang Yang, Scott Yih, Deniz Yuret;

Annie Zaenen, David Zajic, Richard Zens, Luke Zettlemoyer, Dongdong Zhang, Hao Zhang, Tong Zhang, Bing Zhao, Jing Zheng, GuoDong Zhou, Joe Zhou, Jerry Zhu, Imed Zitouni, Onno Zoeter, Andreas Zollmann, Ingrid Zukerman, Geoff Zweig

Table of Contents

Conference Program Overview

Saturday, October 25, 2008

9:15–10:30	Session 1: Plenary Session
10:30–11:00	Morning Break
11:00–12:15	Sessions 2a, 2b and 2c
12:15–14:00	Lunch
14:00–15:15	Sessions 3a, 3b and 3c
15:15–15:45	Afternoon Break
15:45–17:00	Sessions 4a, 4b and 4c
18:00–21:00	Session 5: All Posters

Sunday, October 26, 2008

9:30–10:30	Session 6: Plenary Session
10:30–11:00	Morning Break
11:00–12:15	Sessions 7a, 7b and 7c
12:15–14:00	Lunch
14:00–15:15	Sessions 8a, 8b and 8c
15:15–15:45	Afternoon Break
15:45–17:00	Sessions 9a, 9b and 9c

Monday, October 27, 2008

9:30–10:30	Session 10: Plenary Session
10:30–11:00	Morning Break
11:00–12:15	Sessions 11a, 11b and 11c
12:15-13:00	SIGDAT Business Meeting
13:00–14:00	Lunch
14:00–15:15	Sessions 12a, 12b and 12c
15:15–15:45	Afternoon Break
15:45–17:00	Sessions 13a, 13b and 13c

Conference Program

Saturday, October 25, 2008

Session 1: Plenary Session

9:15–9:30 Opening Remarks

9:30–10:30 Invited Talk: *We KnowItAll: Lessons from a Quarter Century of Web Extraction Research*
Oren Etzioni, University of Washington

10:30-11:00 **Morning Break**

Session 2a: Information Extraction

11:00–11:25 *Revealing the Structure of Medical Dictations with Conditional Random Fields*
Jeremy Jancsary, Johannes Matiasek and Harald Trost

11:25–11:50 *It's a Contradiction – no, it's not: A Case Study using Functional Relations*
Alan Ritter, Stephen Soderland, Doug Downey and Oren Etzioni

11:50–12:15 *Regular Expression Learning for Information Extraction*
Yunyao Li, Rajasekar Krishnamurthy, Sriram Raghavan, Shivakumar Vaithyanathan and H. V. Jagadish

Session 2b: Machine Learning

11:00–11:25 *Modeling Annotators: A Generative Approach to Learning from Annotator Rationales*
Omar Zaidan and Jason Eisner

11:25–11:50 *One-Class Clustering in the Text Domain*
Ron Bekkerman and Koby Crammer

11:50–12:15 *Refining Generative Language Models using Discriminative Learning*
Ben Sandbank

Saturday, October 25, 2008 (continued)

Session 2c: Semantics

11:00–11:25 *Discriminative Learning of Selectional Preference from Unlabeled Text*
Shane Bergsma, Dekang Lin and Randy Goebel

11:25–11:50 *Dependency-based Semantic Role Labeling of PropBank*
Richard Johansson and Pierre Nugues

11:50–12:15 *Scaling Textual Inference to the Web*
Stefan Schoenmackers, Oren Etzioni and Daniel Weld

12:15–14:00 **Lunch**

Session 3a: Machine Translation

14:00–14:25 *Maximum Entropy based Rule Selection Model for Syntax-based Statistical Machine Translation*
Qun Liu, Zhongjun He, Yang Liu and Shouxun Lin

14:25–14:50 *Indirect-HMM-based Hypothesis Alignment for Combining Outputs from Machine Translation Systems*
Xiaodong He, Mei Yang, Jianfeng Gao, Patrick Nguyen and Robert Moore

14:50–15:15 *Coarse-to-Fine Syntactic Machine Translation using Language Projections*
Slav Petrov, Aria Haghighi and Dan Klein

Session 3b: Sentiment Analysis

14:00–14:25 *Adding Redundant Features for CRFs-based Sentence Sentiment Classification*
Jun Zhao, Kang Liu and Gen Wang

14:25–14:50 *Multilingual Subjectivity Analysis Using Machine Translation*
Carmen Banea, Rada Mihalcea, Janyce Wiebe and Samer Hassan

14:50–15:15 *Ranking Reader Emotions Using Pairwise Loss Minimization and Emotional Distribution Regression*
Kevin Hsin-Yih Lin and Hsin-Hsi Chen

Saturday, October 25, 2008 (continued)

Session 3c: Parsing

14:00–14:25 *Dependency Parsing by Belief Propagation*
David Smith and Jason Eisner

14:25–14:50 *Stacking Dependency Parsers*
André Filipe Torres Martins, Dipanjan Das, Noah A. Smith and Eric P. Xing

14:50–15:15 *Better Binarization for the CKY Parsing*
Xinying Song, Shilin Ding and Chin-Yew Lin

15:15–15:45 **Afternoon Break**

Session 4a: Generation

15:45–16:10 *Sentence Fusion via Dependency Graph Compression*
Katja Filippova and Michael Strube

16:10–16:35 *Revisiting Readability: A Unified Framework for Predicting Text Quality*
Emily Pitler and Ani Nenkova

16:35–17:00 *Syntactic Constraints on Paraphrases Extracted from Parallel Corpora*
Chris Callison-Burch

Session 4b: Machine Translation

15:45–16:10 *Forest-based Translation Rule Extraction*
Haitao Mi and Liang Huang

16:10–16:35 *Probabilistic Inference for Machine Translation*
Phil Blunsom and Miles Osborne

16:35–17:00 *Online Large-Margin Training of Syntactic and Structural Translation Features*
David Chiang, Yuval Marton and Philip Resnik

Saturday, October 25, 2008 (continued)

Session 4c: Psycholinguistics

15:45–16:10 *A Noisy-Channel Model of Human Sentence Comprehension under Uncertain Input*
Roger Levy

16:10–16:35 *Incorporating Temporal and Semantic Information with Eye Gaze for Automatic Word Acquisition in Multimodal Conversational Systems*
Shaolin Qu and Joyce Chai

16:35–17:00 *Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks*
Rion Snow, Brendan O'Connor, Daniel Jurafsky and Andrew Ng

18:00-21:00 **Session 5: All Posters**

HotSpots: Visualizing Edits to a Text
Srinivas Bangalore and David Smith

Who is Who and What is What: Experiments in Cross-Document Co-Reference
Alex Baron and Marjorie Freedman

Arabic Named Entity Recognition using Optimized Feature Sets
Yassine Benajiba, Mona Diab and Paolo Rosso

Understanding the Value of Features for Coreference Resolution
Eric Bengtson and Dan Roth

Selecting Sentences for Answering Complex Questions
Yllias Chali and Shafiq Joty

Sampling Alignment Structure under a Bayesian Translation Model
John DeNero, Alexandre Bouchard-Côté and Dan Klein

Improving Chinese Semantic Role Classification with Hierarchical Feature Selection Strategy
Weiwei Ding and Baobao Chang

Bayesian Unsupervised Topic Segmentation
Jacob Eisenstein and Regina Barzilay

Saturday, October 25, 2008 (continued)

A comparison of Bayesian estimators for unsupervised Hidden Markov Model POS taggers
Jianfeng Gao and Mark Johnson

Transliteration as Constrained Optimization
Dan Goldwasser and Dan Roth

Studying the History of Ideas Using Topic Models
David Hall, Daniel Jurafsky and Christopher D. Manning

Triplet Lexicon Models for Statistical Machine Translation
Saša Hasan, Juri Ganitkevitch, Hermann Ney and Jesús Andrés-Ferrer

A Casual Conversation System Using Modality and Word Associations Retrieved from the Web
Shinsuke Higuchi, Rafal Rzepka and Kenji Araki

When Harry Met Harri: Cross-lingual Name Spelling Normalization
Fei Huang, Ahmad Emami and Imed Zitouni

A Dependency-based Word Subsequence Kernel
Rohit Kate

Bridging Lexical Gaps between Queries and Questions on Large Online Q&A Collections with Compact Translation Models
Jung-Tae Lee, Sang-Bum Kim, Young-In Song and Hae-Chang Rim

Scalable Language Processing Algorithms for the Masses: A Case Study in Computing Word Co-occurrence Matrices with MapReduce
Jimmy Lin

Online Acquisition of Japanese Unknown Morphemes using Morphological Constraints
Yugo Murawaki and Sadao Kurohashi

Legal Docket Classification: Where Machine Learning Stumbles
Ramesh Nallapati and Christopher D. Manning

A Discriminative Candidate Generator for String Transformations
Naoaki Okazaki, Yoshimasa Tsuruoka, Sophia Ananiadou and Jun'ichi Tsujii

Saturday, October 25, 2008 (continued)

Automatic induction of FrameNet lexical units
Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce and Michael Roth

Multimodal Subjectivity Analysis of Multiparty Conversation
Stephan Raaijmakers, Khiet Truong and Theresa Wilson

Adapting a Lexicalized-Grammar Parser to Contrasting Domains
Laura Rimell and Stephen Clark

Improving Interactive Machine Translation via Mouse Actions
Germán Sanchis-Trilles, Daniel Ortiz-Martínez, Jorge Civera, Francisco Casacuberta, Enrique Vidal and Hieu Hoang

LTAG Dependency Parsing with Bidirectional Incremental Construction
Libin Shen and Aravind Joshi

Improved Sentence Alignment on Parallel Web Pages Using a Stochastic Tree Alignment Model
Lei Shi and Ming Zhou

HTM: A Topic Model for Hypertexts
Congkai Sun, Bin Gao, Zhenfu Cao and Hang Li

A Japanese Predicate Argument Structure Analysis using Decision Lists
Hirotoshi Taira, Sanae Fujita and Masaaki Nagata

Online Word Games for Semantic Data Collection
David Vickrey, Aaron Bronzan, William Choi, Aman Kumar, Jason Turner-Maier, Arthur Wang and Daphne Koller

Seed and Grow: Augmenting Statistically Generated Summary Sentences using Schematic Word Patterns
Stephen Wan, Robert Dale, Mark Dras and Cecile Paris

Using Bilingual Knowledge and Ensemble Techniques for Unsupervised Chinese Sentiment Analysis
Xiaojun Wan

A Tale of Two Parsers: Investigating and Combining Graph-based and Transition-based Dependency Parsing
Yue Zhang and Stephen Clark

Saturday, October 25, 2008 (continued)

Generalizing Local and Non-Local Word-Reordering Patterns for Syntax-Based Machine Translation
Bing Zhao and Yaser Al-Onaizan

Sunday, October 26, 2008

Session 6: Plenary Session

9:30–10:30 Invited Talk: *Connecting language learning and language evolution via Bayesian statistics*
Tom Griffiths, University of California, Berkeley

10:30-11:00 **Morning Break**

Session 7a: Information Extraction

11:00–11:25 *Weakly-Supervised Acquisition of Labeled Class Instances using Graph Random Walks*
Partha Pratim Talukdar, Joseph Reisinger, Marius Pasca, Deepak Ravichandran, Rahul Bhagat and Fernando Pereira

11:25–11:50 *Seeded Discovery of Base Relations in Large Corpora*
Nicholas Andrews and Naren Ramakrishnan

11:50–12:15 *Mention Detection Crossing the Language Barrier*
Imed Zitouni and Radu Florian

Session 7b: Machine Translation

11:00–11:25 *Decomposability of Translation Metrics for Improved Evaluation and Efficient Algorithms*
David Chiang, Steve DeNeefe, Yee Seng Chan and Hwee Tou Ng

11:25–11:50 *Lattice Minimum Bayes-Risk Decoding for Statistical Machine Translation*
Roy Tromble, Shankar Kumar, Franz Och and Wolfgang Macherey

11:50–12:15 *Phrase Translation Probabilities with ITG Priors and Smoothing as Learning Objective*
Markos Mylonakis and Khalil Sima'an

Sunday, October 26, 2008 (continued)

Session 7c: Coreference Resolution

11:00–11:25 *Unsupervised Models for Coreference Resolution*
Vincent Ng

11:25–11:50 *Joint Unsupervised Coreference Resolution with Markov Logic*
Hoifung Poon and Pedro Domingos

11:50–12:15 *Specialized Models and Ranking for Coreference Resolution*
Pascal Denis and Jason Baldridge

12:15–14:00 **Lunch**

Session 8a: Machine Learning

14:00–14:25 *Learning with Probabilistic Features for Improved Pipeline Models*
Razvan Bunescu

14:25–14:50 *Cross-Task Knowledge-Constrained Self Training*
Hal Daumé III

14:50–15:15 *Online Methods for Multi-Domain Learning and Adaptation*
Mark Dredze and Koby Crammer

Session 8b: Semantics

14:00–14:25 *Jointly Combining Implicit Constraints Improves Temporal Ordering*
Nathanael Chambers and Daniel Jurafsky

14:25–14:50 *Automatic Inference of the Temporal Location of Situations in Chinese Text*
Nianwen Xue

14:50–15:15 *Learning the Scope of Negation in Biomedical Texts*
Roser Morante, Anthony Liekens and Walter Daelemans

Sunday, October 26, 2008 (continued)

Session 8c: Machine Translation

14:00–14:25 *Lattice-based Minimum Error Rate Training for Statistical Machine Translation*
Wolfgang Macherey, Franz Och, Ignacio Thayer and Jakob Uszkoreit

14:25–14:50 *Syntactic Models for Structural Word Insertion and Deletion during Translation*
Arul Menezes and Chris Quirk

14:50–15:15 *Predicting Success in Machine Translation*
Alexandra Birch, Miles Osborne and Philipp Koehn

15:15–15:45 **Afternoon Break**

Session 9a: Summarization

15:45–16:10 *An Exploration of Document Impact on Graph-Based Multi-Document Summarization*
Xiaojun Wan

16:10–16:35 *Topic-Driven Multi-Document Summarization with Encyclopedic Knowledge and Spreading Activation*
Vivi Nastase

16:35–17:00 *Summarizing Spoken and Written Conversations*
Gabriel Murray and Giuseppe Carenini

Session 9b: Semantics

15:45–16:10 *A Generative Model for Parsing Natural Language to Meaning Representations*
Wei Lu, Hwee Tou Ng, Wee Sun Lee and Luke S. Zettlemoyer

16:10–16:35 *Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis*
Yejin Choi and Claire Cardie

16:35–17:00 *A Phrase-Based Alignment Model for Natural Language Inference*
Bill MacCartney, Michel Galley and Christopher D. Manning

Sunday, October 26, 2008 (continued)

Session 9c: Language Modeling

15:45–16:10 *Attacking Decipherment Problems Optimally with Low-Order N-gram Models*
Sujith Ravi and Kevin Knight

16:10–16:35 *Integrating Multi-level Linguistic Knowledge with a Unified Framework for Mandarin Speech Recognition*
Xinhao Wang, Jiazhong Nie, Dingsheng Luo and Xihong Wu

16:35–17:00 *N-gram Weighting: Reducing Training Data Mismatch in Cross-Domain Language Model Estimation*
Bo-June (Paul) Hsu and James Glass

Monday, October 27, 2008

Session 10: Plenary Session

9:30–10:30 Invited Talk: *Are Linear Models Right for Language?*
Fernando Pereira, Google and University of Pennsylvania

10:30-11:00 **Morning Break**

Session 11a: Machine Translation

11:00–11:25 *Complexity of Finding the BLEU-optimal Hypothesis in a Confusion Network*
Gregor Leusch, Evgeny Matusov and Hermann Ney

11:25–11:50 *A Simple and Effective Hierarchical Phrase Reordering Model*
Michel Galley and Christopher D. Manning

11:50–12:15 *Language and Translation Model Adaptation using Comparable Corpora*
Matthew Snover, Bonnie Dorr and Richard Schwartz

Monday, October 27, 2008 (continued)

Session 11b: Parsing

11:00–11:25 *Sparse Multi-Scale Grammars for Discriminative Latent Variable Parsing*
Slav Petrov and Dan Klein

11:25–11:50 *Two Languages are Better than One (for Syntactic Parsing)*
David Burkett and Dan Klein

11:50–12:15 *Automatic Prediction of Parser Accuracy*
Sujith Ravi, Kevin Knight and Radu Soricut

Session 11c: Semantics

11:00–11:25 *A Structured Vector Space Model for Word Meaning in Context*
Katrin Erk and Sebastian Padó

11:25–11:50 *Learning Graph Walk Based Similarity Measures for Parsed Text*
Einat Minkov and William W. Cohen

11:50–12:15 *A Graph-theoretic Model of Lexical Syntactic Acquisition*
Hinrich Schütze and Michael Walsh

12:15–13:00 **SIGDAT Business Meeting**

13:00–14:00 **Lunch**

Monday, October 27, 2008 (continued)

Session 12a: Question Answering

14:00–14:25 *Question Classification using Head Words and their Hypernyms*
Zhiheng Huang, Marcus Thint and Zengchang Qin

14:25–14:50 *CoCQA: Co-Training over Questions and Answers with an Application to Predicting Question Subjectivity Orientation*
Baoli Li, Yandong Liu and Eugene Agichtein

14:50–15:15 *Automatic Set Expansion for List Question Answering*
Richard C. Wang, Nico Schlaefer, William W. Cohen and Eric Nyberg

Session 12b: Dialogue

14:00–14:25 *Acquiring Domain-Specific Dialog Information from Task-Oriented Human-Human Interaction through an Unsupervised Learning*
Ananlada Chotimongkol and Alexander Rudnicky

14:25–14:50 *Relative Rank Statistics for Dialog Analysis*
Juan Huerta

14:50–15:15 *Learning to Predict Code-Switching Points*
Thamar Solorio and Yang Liu

Session 12c: Semantics

14:00–14:25 *Computing Word-Pair Antonymy*
Saif Mohammad, Bonnie Dorr and Graeme Hirst

14:25–14:50 *Construction of an Idiom Corpus and its Application to Idiom Identification based on WSD Incorporating Idiom-Specific Features*
Chikara Hashimoto and Daisuke Kawahara

14:50–15:15 *Word Sense Disambiguation Using OntoNotes: An Empirical Study*
Zhi Zhong, Hwee Tou Ng and Yee Seng Chan

15:15–15:45 **Afternoon Break**

Monday, October 27, 2008 (continued)

Session 13a: Information Extraction

15:45–16:10 *Graph-based Analysis of Semantic Drift in Espresso-like Bootstrapping Algorithms*
Mamoru Komachi, Taku Kudo, Masashi Shimbo and Yuji Matsumoto

16:10–16:35 *The Linguistic Structure of English Web-Search Queries*
Cory Barr, Rosie Jones and Moira Regelson

16:35–17:00 *Mining and Modeling Relations between Formal and Informal Chinese Phrases from Web Corpora*
Zhifei Li and David Yarowsky

Session 13b: POS tagging and Word Segmentation

15:45–16:10 *Unsupervised Multilingual Learning for POS Tagging*
Benjamin Snyder, Tahira Naseem, Jacob Eisenstein and Regina Barzilay

16:10–16:35 *Part-of-Speech Tagging for English-Spanish Code-Switched Text*
Thamar Solorio and Yang Liu

16:35–17:00 *Information Retrieval Oriented Word Segmentation based on Character Association Strength Ranking*
Yixuan Liu, Bin Wang, Fan Ding and Sheng Xu

Session 13c: Machine Learning

15:45–16:10 *An Analysis of Active Learning Strategies for Sequence Labeling Tasks*
Burr Settles and Mark Craven

16:10–16:35 *Latent-Variable Modeling of String Transductions with Finite-State Methods*
Markus Dreyer, Jason Smith and Jason Eisner

16:35–17:00 *Soft-Supervised Learning for Text Classification*
Amarnag Subramanya and Jeff Bilmes