EMNLP 2009

Proceedings of the 2009 Conference on

Empirical Methods in Natural Language Processing

A meeting of SIGDAT,
a special interest group of ACL
held in conjunction with ACL-IJCNLP 2009

6–7 August 2009
Singapore

Production and Manufacturing by
World Scientific Publishing Co Pte Ltd
5 Toh Tuck Link
Singapore 596224

The conference organizers are grateful to Microsoft Research for their generous support.

Microsoft®
Research

©2009 The Association for Computational Linguistics
and The Asian Federation of Natural Language Processing

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 978-1-932432-59-6 / 1-932432-59-0 (Volume 1)
ISBN 978-1-932432-62-6 / 1-932432-62-0 (Volume 2)
ISBN 978-1-932432-63-3 / 1-932432-63-9 (Volume 3)

Introduction

Welcome to the 2009 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 ACL-IJCNLP 2009 in Singapore.

EMNLP received 475 submissions, a new record. We were able to accept 163 papers in total (an acceptance rate of 34%). Of these, 96 (20%) were accepted for oral presentation, and 67 (14%) for poster presentation. The papers were selected by a program committee of 15 area chairs, from Asia, Europe, and North America, assisted by a panel of 389 reviewers. This year EMNLP again held 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. Additional thanks to the Publications Chair, David Chiang, 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."[1] Special thanks to David Yarowsky and Ken Church of SIGDAT who provided much valuable advice and assistance over the past months. We are akso grateful for the financial support from Microsoft.

We are most grateful to Haizhou Li 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 predecessors, Mirella Lapata and Hwee Tou Ng, whose experience and example we shamelessly exploited.

Philipp Koehn
Rada Mihalcea

[1] http://www.cs.jhu.edu/jason/advice/how-to-chair-a-conference.html

Program Co-Chairs:

Philipp Koehn, University of Edinburgh
Rada Mihalcea, University of North Texas

Area Chairs:

Stephen Clark, University of Cambridge
Mona Diab, Columbia University
Jason Eisner, Johns Hopkins University
Katrin Erk, University of Texas
Eric Fosler-Lussier, Ohio State University
Iryna Gurevych, Darmstadt University
Hang Li, Microsoft
Chin-Yew Lin, Microsoft
Adam Lopez, University of Edinburgh
Vivi Năstase, EML Research
Miles Osborne, University of Edinburgh
Tim Paek, Microsoft
Marius Paşca, Google
Carlo Strapparava, FBK-irst
Theresa Wilson, University of Edinburgh

Local Arrangements Chair:

Haizhou Li, Institute for Infocomm Research

Publications Chair:

David Chiang, USC Information Sciences Institute

Reviewers:

Ahmed Abbasi
Eugene Agichtein
Amr Ahmed
Ben Allison
Yasemin Altun
Alina Andreevskaia
Shlomo Argamon
Abhishek Arun
Jordi Atserias
Necip Fazil Ayan
Timothy Baldwim
Srinivas Bangalore
Roy Bar-Haim
Regina Barzilay
Roberto Basili
Sabine Bergler
Steven Bethard
Rahul Bhagat
Chris Biemann
Alexandra Birch
Phil Blunsom
Dan Bohus
Gemma Boleda
Johan Bos
Alexandre Bouchard-Côté
Thorsten Brants
Eric Breck
Sam Brody
Susan Brown
Paul Buitelaar
Razvan Bunescu
Aljoscha Burchardt
Jill Burstein
Bill Byrne
Chris Callison-Burch
Nicola Cancedda
Yunbo Cao
Guiseppe Carenini
Marine Carpuat
Xavier Carreras
John Carroll
Vitor Carvalho
Francisco Casacuberta
Mauro Cettolo
Nate Chambers
Yee Seng Chan
Pi-Chuan Chang
Harr Chen
Hsin-Hsi Chen
Pu-Jen Cheng
Xueqi Cheng
Colin Cherry
David Chiang
Yejin Choi
Jennifer Chu-Carroll
Ken Church
Massimiliano Ciaramita
James Clarke
Paul Cook
Bonaventura Coppola
Marta R. Costa-Jussa
Mathias Creutz
Montse Cuadros Oller
Silviu Cucerzan
Hang Cui
James Curran
Walter Daelemans
Hercules Dalianis
Dipanjan Das
Dmitry Davidov
Adrià de Gispert
Marie-Catherine de Marneffe
Steve DeNeefe
John DeNero
Yonggang Deng
Pascal Denis
Ann Devitt
Fernando Diaz
Markus Dickinson
Mark Dredze
Markus Dreyer
Amit Dubey
Chris Dyer
Koji Eguchi
Andreas Eisele
Michael Elhadad
Andrea Esuli
Stefan Evert
Richard Farkas
Afsaneh Fazly
Christiane Fellbaum
Jenny Finkel
Margaret Fleck
Radu Florian
George Foster
Anette Frank
Alex Fraser
Pascale Fung
Evgeniy Gabrilovich
Michel Galley
Michael Gamon
Jianfeng Gao
Yuqing Gao
Nikesh Garera
Kallirroi Georgila
Ulrich Germann
Daniel Gildea
Jesús Giménez
Jonathan Ginzburg
Roxana Girju
Claudio Giuliano
Jim Glass
Alfio Gliozzo
Andrew Goldberg
John Goldsmith
Sharon Goldwater
Carlos Gomez-Rodriguez
Julio Gonzalo
Cyril Goutte
Mark Greenwood
Nizar Habash
Barry Haddow
Patrick Haffner
Thomas Hain
Dilek Hakkani-Tur
David Hall
Keith Hall
Sanda Harabagiu
Saša Hasan
Mark Hasegawa-Johnson
Xiaodong He
Jeffrey Heinz
James Henderson
John Henderson
Graeme Hirst
Hieu Hoang
Julia Hockenmaier
Kristy Hollingshead
Tracy Holloway King
Florentina Hristea
Liang Huang
Lluis Hurtado
Nancy Ide
Gonzalo Iglesias
Diana Inkpen
Martin Jansche
Nitin Jindal
Richard Johansson
Pamela Jordan
Joemon Jose
Hiroshi Kanayama
Rohit Kate
Graham Katz
Tatsuya Kawahara
Frank Keller
Adam Kilgarriff
Soo-Min Kim
Dietrich Klakow
Alexandre Klementiev
Kevin Knight
Rob Koeling
Grzegorz Kondrak
Terry Koo
Moshe Koppel
András Kornai
Zornitsa Kozareva
Lun-Wei Ku
Sandra Kuebler
Roland Kuhn
Ravi Kumar
Shankar Kumar
Sadao Kurohashi
Wei Lai
Wai Lam
Lori Lamel
Alon Lavie
Victor Lavrenko
Matthew Lease
Gary Lee
Gina-Anne Levow
Wei Li
Zhifei Li
Dekang Lin
Hsuan-Tien Lin
Jimmy Lin
Shou-de Lin
Lucian Lita
Bing Liu
Nathan Liu
Qun Liu
Yang Liu

Karen Livescu
Wolfgang Macherey
Nitin Madnani
Rob Malouf
Gideon Mann
Chris Manning
Daniel Marcu
Katja Markert
David Martinez
Andre Martins
Yuji Matsumoto
Takuya Matsuzaki
Mark Maybury
Andrew McCallum
Diana McCarthy
Ryan McDonald
Kathy McKeown
Qiaozhu Mei
Arul Menezes
Florian Metze
Donald Metzler
Haitao Mi
Einat Minkov
Shachar Mirkin
Jeff Mitchell
Vibhu Mittal
Yusuke Miyao
Marie-Francine Moens
Saif Mohammad
Christof Monz
Tsun Moon
Raymond Mooney
Robert Moore
Alessandro Moschitti
Tony Mullen
Smaranda Muresan
Gabriel Murray
Gabriele Musillo
Sung-Hyon Myaeng
Mikio Nakano
Preslav Nakov
Su Nam Kim
Alexis Nasr
Tetsuya Nasukawa
Roberto Navigli
Mark-Jan Nederhof
Ani Nenkova
Gunther Neumann
Hwee Tou Ng
Vincent Ng
Nicolas Nicolov
Rodney Nielsen
Joakim Nivre
Tadashi Nomoto
Franz Och
Kemal Oflazer
Manubu Okumura
Jahna Otterbacher
Iadh Ounis
Ulrike Pado
Lluis Padro
Bo Pang
Patrick Pantel
Fuchun Peng
Marco Pennacchiotti
Slav Petrov
Emanuele Pianta
Olivier Pietquin
Daniele Pighin
Massimo Poesio
Elias Ponvert
Simone Ponzetto
Ana-Maria Popescu
Marius Popescu
Maja Popović
Victor Poznanski
John Prager
Rashmi Prasad
Partha Pratim Talukdar
Matthew Purver
Silvia Quarteroni
Chris Quirk
Stephan Raaijmakers
Dan Ramage
Owen Rambow
Delip Rao
Ari Rappoport
Deepak Ravichandran
Roi Reichart
Sebastian Riedel
Stefan Riezler
German Rigau
Michael Riley
Hae-Chang Rim
Laura Rimell
Brian Roark
Paolo Rosso
Antti-Veikko Rosti
Alex Rudnicky
Anna Rumshisky
Vasile Rus
Anton Rytting
Marta Sabou
Bogdan Sacaleanu
Kenji Sagae
Horacio Saggion
Murat Saraclar
Anoop Sarkar
Roser Sauri
Helmut Schmid
Karin Schuler
Sabine Schulte im Walde
Holger Schwenk
Fabrizio Sebastiani
Frederique Segond
Yohei Seki
Satoshi Sekine
Stephanie Seneff
Hendra Setiawan
Burr Settles
Izhak Shafran
Dou Shen
Khalil Sima'an
Michel Simard
David Smith
Noah Smith
Matthew Snover
Swapna Somasundaran
Young-In Song
Aitor Soroa
Caroline Sporleder
Richard Sproat
Rohini Srihari
Mark Steedman
Amanda Stent
Mark Stevenson
Veselin Stoyanov
Michael Strube
Eiichiro Sumita
Maosong Sun
Mihai Surdeanu
Gyrgy Szarvas
Stan Szpakowicz
Idan Szpektor
Hiroya Takamura
David Talbot
Joel Tetreault
Simone Teufel
Christoph Tillmann
Ivan Titov
Kristina Toutanova
Roy Tromble
Yoshimasa Tsuruoka
Antal van den Bosch
Gertjan van Noord
Josef van Genabith
Lonneke van der Plas
Benjamin Van Durme
Lucy Vanderwende
Sebastian Varges
Ashish Venugopal
David Vilar
Stephan Vogel
Piek Vossen
Kuansan Wang
Richard Wang
Wei Wang
Taro Watanabe
Andy Way
Amy Weinberg
David Weir
Dan Weld
Michael White
Richard Wicentowski
Jan Wiebe
Jason Williams
Shuly Wintner
Rene Witte
Wensi Xi
Fei Xia
Deyi Xiong
Gu Xu
Jun Xu
Deniz Yuret
Richard Zens
Dell Zhang
Hao Zhang
Min Zhang
Yue Zhang
Bing Zhao
Min Zhao
Tie-Jun Zhao
Ming Zhou
Shenghuo Zhu
Imed Zitouni
Onno Zoeter
Andreas Zollmann
Ingrid Zukerman

Table of Contents

Unsupervised Semantic Parsing
Hoifung Poon and Pedro Domingos 1

Graph Alignment for Semi-Supervised Semantic Role Labeling
Hagen Fürstenau and Mirella Lapata 11

Semi-supervised Semantic Role Labeling Using the Latent Words Language Model
Koen Deschacht and Marie-Francine Moens 21

Semantic Dependency Parsing of NomBank and PropBank: An Efficient Integrated Approach via a Large-scale Feature Selection
Hai Zhao, Wenliang Chen and Chunyu Kit 30

First- and Second-Order Expectation Semirings with Applications to Minimum-Risk Training on Translation Forests
Zhifei Li and Jason Eisner 40

Feasibility of Human-in-the-loop Minimum Error Rate Training
Omar F. Zaidan and Chris Callison-Burch 52

Cube Pruning as Heuristic Search
Mark Hopkins and Greg Langmead 62

Effective Use of Linguistic and Contextual Information for Statistical Machine Translation
Libin Shen, Jinxi Xu, Bing Zhang, Spyros Matsoukas and Ralph Weischedel 72

Active Learning by Labeling Features
Gregory Druck, Burr Settles and Andrew McCallum 81

Efficient Kernels for Sentence Pair Classification
Fabio Massimo Zanzotto and Lorenzo Dell'Arciprete 91

Graphical Models over Multiple Strings
Markus Dreyer and Jason Eisner 101

Reverse Engineering of Tree Kernel Feature Spaces
Daniele Pighin and Alessandro Moschitti 111

A Rich Feature Vector for Protein-Protein Interaction Extraction from Multiple Corpora
Makoto Miwa, Rune Sætre, Yusuke Miyao and Jun'ichi Tsujii 121

Generalized Expectation Criteria for Bootstrapping Extractors using Record-Text Alignment
Kedar Bellare and Andrew McCallum 131

Nested Named Entity Recognition
Jenny Rose Finkel and Christopher D. Manning 141

A Unified Model of Phrasal and Sentential Evidence for Information Extraction
Siddharth Patwardhan and Ellen Riloff 151

Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm
Jingjing Liu and Stephanie Seneff 161

Supervised and Unsupervised Methods in Employing Discourse Relations for Improving Opinion Polarity Classification
Swapna Somasundaran, Galileo Namata, Janyce Wiebe and Lise Getoor 170

Sentiment Analysis of Conditional Sentences
Ramanathan Narayanan, Bing Liu and Alok Choudhary 180

Subjectivity Word Sense Disambiguation
Cem Akkaya, Janyce Wiebe and Rada Mihalcea 190

Non-Projective Parsing for Statistical Machine Translation
Xavier Carreras and Michael Collins 200

Extending Statistical Machine Translation with Discriminative and Trigger-Based Lexicon Models
Arne Mauser, Saša Hasan and Hermann Ney 210

Feature-Rich Translation by Quasi-Synchronous Lattice Parsing
Kevin Gimpel and Noah A. Smith 219

Improved Word Alignment with Statistics and Linguistic Heuristics
Ulf Hermjakob 229

Entity Extraction via Ensemble Semantics
Marco Pennacchiotti and Patrick Pantel 238

Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora
Daniel Ramage, David Hall, Ramesh Nallapati and Christopher D. Manning 248

Clustering to Find Exemplar Terms for Keyphrase Extraction
Zhiyuan Liu, Peng Li, Yabin Zheng and Maosong Sun 257

Geo-mining: Discovery of Road and Transport Networks Using Directional Patterns
Dmitry Davidov and Ari Rappoport 267

Wikipedia as Frame Information Repository
Sara Tonelli and Claudio Giuliano 276

Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk
Chris Callison-Burch 286

How Well Does Active Learning Actually *Work? Time-based Evaluation of Cost-reduction Strategies for Language Documentation*
Jason Baldridge and Alexis Palmer 296

Automatically Evaluating Content Selection in Summarization without Human Models
Annie Louis and Ani Nenkova 306

Classifier Combination for Contextual Idiom Detection Without Labelled Data
Linlin Li and Caroline Sporleder 315

Deriving Lexical and Syntactic Expectation-based Measures for Psycholinguistic Modeling via Incremental Top-down Parsing
Brian Roark, Asaf Bachrach, Carlos Cardenas and Christophe Pallier 324

It's Not You, it's Me: Detecting Flirting and its Misperception in Speed-Dates
Rajesh Ranganath, Dan Jurafsky and Dan McFarland 334

Recognizing Implicit Discourse Relations in the Penn Discourse Treebank
Ziheng Lin, Min-Yen Kan and Hwee Tou Ng 343

A Bayesian Model of Syntax-Directed Tree to String Grammar Induction
Trevor Cohn and Phil Blunsom 352

Better Synchronous Binarization for Machine Translation
Tong Xiao, Mu Li, Dongdong Zhang, Jingbo Zhu and Ming Zhou 362

Accuracy-Based Scoring for DOT: Towards Direct Error Minimization for Data-Oriented Translation
Daniel Galron, Sergio Penkale, Andy Way and I. Dan Melamed 371

Improved Statistical Machine Translation Using Monolingually-Derived Paraphrases
Yuval Marton, Chris Callison-Burch and Philip Resnik 381

A Comparison of Model Free versus Model Intensive Approaches to Sentence Compression
Tadashi Nomoto 391

Natural Language Generation with Tree Conditional Random Fields
Wei Lu, Hwee Tou Ng and Wee Sun Lee 400

Perceptron Reranking for CCG Realization
Michael White and Rajakrishnan Rajkumar 410

Multi-Document Summarisation Using Generic Relation Extraction
Ben Hachey 420

Language Models Based on Semantic Composition
Jeff Mitchell and Mirella Lapata 430

Graded Word Sense Assignment
Katrin Erk and Diana McCarthy 440

Joint Learning of Preposition Senses and Semantic Roles of Prepositional Phrases
Daniel Dahlmeier, Hwee Tou Ng and Tanja Schultz 450

Projecting Parameters for Multilingual Word Sense Disambiguation
Mitesh M. Khapra, Sapan Shah, Piyush Kedia and Pushpak Bhattacharyya 459

Multi-Word Expression Identification Using Sentence Surface Features
Ram Boukobza and Ari Rappoport 468

Acquiring Translation Equivalences of Multiword Expressions by Normalized Correlation Frequencies
Ming-Hong Bai, Jia-Ming You, Keh-Jiann Chen and Jason S. Chang 478

Collocation Extraction Using Monolingual Word Alignment Method
Zhanyi Liu, Haifeng Wang, Hua Wu and Sheng Li 487

Multi-Class Confidence Weighted Algorithms
Koby Crammer, Mark Dredze and Alex Kulesza 496

Model Adaptation via Model Interpolation and Boosting for Web Search Ranking
Jianfeng Gao, Qiang Wu, Chris Burges, Krysta Svore, Yi Su, Nazan Khan, Shalin Shah and Hongyan Zhou 505

A Structural Support Vector Method for Extracting Contexts and Answers of Questions from Online Forums
Wen-Yun Yang, Yunbo Cao and Chin-Yew Lin 514

Mining Search Engine Clickthrough Log for Matching N-gram Features
Huihsin Tseng, Longbin Chen, Fan Li, Ziming Zhuang, Lei Duan and Belle Tseng 524

The Role of Named Entities in Web People Search
Javier Artiles, Enrique Amigó and Julio Gonzalo 534

Investigation of Question Classifier in Question Answering
Zhiheng Huang, Marcus Thint and Asli Celikyilmaz 543

An Empirical Study of Semi-supervised Structured Conditional Models for Dependency Parsing
Jun Suzuki, Hideki Isozaki, Xavier Carreras and Michael Collins 551

Statistical Bistratal Dependency Parsing
Richard Johansson 561

Improving Dependency Parsing with Subtrees from Auto-Parsed Data
Wenliang Chen, Jun'ichi Kazama, Kiyotaka Uchimoto and Kentaro Torisawa 570

Topic-wise, Sentiment-wise, or Otherwise? Identifying the Hidden Dimension for Unsupervised Text Classification
Sajib Dasgupta and Vincent Ng 580

Adapting a Polarity Lexicon using Integer Linear Programming for Domain-Specific Sentiment Classification
Yejin Choi and Claire Cardie 590

Generating High-Coverage Semantic Orientation Lexicons From Overtly Marked Words and a Thesaurus
Saif Mohammad, Cody Dunne and Bonnie Dorr 599

Matching Reviews to Objects using a Language Model
Nilesh Dalvi, Ravi Kumar, Bo Pang and Andrew Tomkins 609

EEG Responds to Conceptual Stimuli and Corpus Semantics
Brian Murphy, Marco Baroni and Massimo Poesio 619

A Comparison of Windowless and Window-Based Computational Association Measures as Predictors of Syntagmatic Human Associations
Justin Washtell and Katja Markert 628

Improving Verb Clustering with Automatically Acquired Selectional Preferences
Lin Sun and Anna Korhonen 638

Improving Web Search Relevance with Semantic Features
Yumao Lu, Fuchun Peng, Gilad Mishne, Xing Wei and Benoit Dumoulin 648

Can Chinese Phonemes Improve Machine Transliteration?: A Comparative Study of English-to-Chinese Transliteration Models
Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa 658

Unsupervised Morphological Segmentation and Clustering with Document Boundaries
Taesun Moon, Katrin Erk and Jason Baldridge 668

The Infinite HMM for Unsupervised PoS Tagging
Jurgen Van Gael, Andreas Vlachos and Zoubin Ghahramani 678

A Simple Unsupervised Learner for POS Disambiguation Rules Given Only a Minimal Lexicon
Qiuye Zhao and Mitch Marcus 688

Tree Kernel-based SVM with Structured Syntactic Knowledge for BTG-based Phrase Reordering
Min Zhang and Haizhou Li 698

Discriminative Corpus Weight Estimation for Machine Translation
Spyros Matsoukas, Antti-Veikko I. Rosti and Bing Zhang 708

Unsupervised Tokenization for Machine Translation
Tagyoung Chung and Daniel Gildea 718

Synchronous Tree Adjoining Machine Translation
Steve DeNeefe and Kevin Knight 727

Word Buffering Models for Improved Speech Repair Parsing
Tim Miller 737

Less is More: Significance-Based N-gram Selection for Smaller, Better Language Models
Robert C. Moore and Chris Quirk 746

Stream-based Randomised Language Models for SMT
Abby Levenberg and Miles Osborne 756

Integrating Sentence- and Word-level Error Identification for Disfluency Correction
Erin Fitzgerald, Frederick Jelinek and Keith Hall 765

Estimating Semantic Distance Using Soft Semantic Constraints in Knowledge-Source – Corpus Hybrid Models
Yuval Marton, Saif Mohammad and Philip Resnik 775

Recognizing Textual Relatedness with Predicate-Argument Structures
Rui Wang and Yi Zhang 784

Learning Term-weighting Functions for Similarity Measures
Wen-tau Yih 793

A Relational Model of Semantic Similarity between Words using Automatically Extracted Lexical Pattern Clusters from the Web
Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka 803

Unbounded Dependency Recovery for Parser Evaluation
Laura Rimell, Stephen Clark and Mark Steedman 813

Parser Adaptation and Projection with Quasi-Synchronous Grammar Features
David A. Smith and Jason Eisner 822

Self-Training PCFG Grammars with Latent Annotations Across Languages
Zhongqiang Huang and Mary Harper 832

An Alternative to Head-Driven Approaches for Parsing a (Relatively) Free Word-Order Language
Reut Tsarfaty, Khalil Sima'an and Remko Scha 842

Enhancement of Lexical Concepts Using Cross-lingual Web Mining
Dmitry Davidov and Ari Rappoport 852

Bilingual Dictionary Generation for Low-resourced Language Pairs
István Varga and Shoichi Yokoyama 862

Multilingual Spectral Clustering Using Document Similarity Propagation
Dani Yogatama and Kumiko Tanaka-Ishii 871

Polylingual Topic Models
David Mimno, Hanna M. Wallach, Jason Naradowsky, David A. Smith & Andrew McCallum . 880

Using the Web for Language Independent Spellchecking and Autocorrection
Casey Whitelaw, Ben Hutchinson, Grace Y Chung and Ged Ellis 890

Statistical Estimation of Word Acquisition with Application to Readability Prediction
Paul Kidwell, Guy Lebanon and Kevyn Collins-Thompson 900

Combining Collocations, Lexical and Encyclopedic Knowledge for Metonymy Resolution
Vivi Nastase and Michael Strube 910

Segmenting Email Message Text into Zones
Andrew Lampert, Robert Dale and Cécile Paris 919

Hypernym Discovery Based on Distributional Similarity and Hierarchical Structures
Ichiro Yamada, Kentaro Torisawa, Jun'ichi Kazama, Kow Kuroda, Masaki Murata, Stijn De Saeger, Francis Bond and Asuka Sumida 929

Web-Scale Distributional Similarity and Entity Set Expansion
Patrick Pantel, Eric Crestan, Arkady Borkovsky, Ana-Maria Popescu and Vishnu Vyas 938

Toward Completeness in Concept Extraction and Classification
Eduard Hovy, Zornitsa Kozareva and Ellen Riloff 948

Reading to Learn: Constructing Features from Semantic Abstracts
Jacob Eisenstein, James Clarke, Dan Goldwasser and Dan Roth 958

Supervised Models for Coreference Resolution
Altaf Rahman and Vincent Ng 968

Global Learning of Noun Phrase Anaphoricity in Coreference Resolution via Label Propagation
GuoDong Zhou and Fang Kong 978

Employing the Centering Theory in Pronoun Resolution from the Semantic Perspective
Fang Kong, GuoDong Zhou and Qiaoming Zhu 987

Person Cross Document Coreference with Name Perplexity Estimates
Octavian Popescu 997

Learning Linear Ordering Problems for Better Translation
Roy Tromble and Jason Eisner 1007

Weighted Alignment Matrices for Statistical Machine Translation
Yang Liu, Tian Xia, Xinyan Xiao and Qun Liu 1017

Sinuhe – Statistical Machine Translation using a Globally Trained Conditional Exponential Family Translation Model
Matti Kääriäinen 1027

Fast Translation Rule Matching for Syntax-based Statistical Machine Translation
Hui Zhang, Min Zhang, Haizhou Li and Chew Lim Tan 1037

Gazpacho and Summer Rash: Lexical Relationships from Temporal Patterns of Web Search Queries
Enrique Alfonseca, Massimiliano Ciaramita and Keith Hall 1046

A Compact Forest for Scalable Inference over Entailment and Paraphrase Rules
Roy Bar-Haim, Jonathan Berant and Ido Dagan 1056

Discriminative Substring Decoding for Transliteration
Colin Cherry and Hisami Suzuki 1066

Re-Ranking Models Based-on Small Training Data for Spoken Language Understanding
Marco Dinarelli, Alessandro Moschitti and Giuseppe Riccardi 1076

Empirical Exploitation of Click Data for Task Specific Ranking
Anlei Dong, Yi Chang, Shihao Ji, Ciya Liao, Xin Li and Zhaohui Zheng 1086

The Feature Subspace Method for SMT System Combination
Nan Duan, Mu Li, Tong Xiao and Ming Zhou 1096

Lattice-based System Combination for Statistical Machine Translation
Yang Feng, Yang Liu, Haitao Mi, Qun Liu and Yajuan Lü 1105

A Joint Language Model With Fine-grain Syntactic Tags
Denis Filimonov and Mary Harper 1114

Bidirectional Phrase-based Statistical Machine Translation
Andrew Finch and Eiichiro Sumita 1124

Real-time Decision Detection in Multi-party Dialogue
Matthew Frampton, Jia Huang, Trung Bui and Stanley Peters 1133

On the Role of Lexical Features in Sequence Labeling
Yoav Goldberg and Michael Elhadad 1142

Simple Coreference Resolution with Rich Syntactic and Semantic Features
Aria Haghighi and Dan Klein 1152

Descriptive and Empirical Approaches to Capturing Underlying Dependencies among Parsing Errors
Tadayoshi Hara, Yusuke Miyao and Jun'ichi Tsujii 1162

Large-Scale Verb Entailment Acquisition from the Web
Chikara Hashimoto, Kentaro Torisawa, Kow Kuroda, Stijn De Saeger, Masaki Murata and Jun'ichi Kazama 1172

A Syntactified Direct Translation Model with Linear-time Decoding
Hany Hassan, Khalil Sima'an and Andy Way 1182

Cross-lingual Semantic Relatedness Using Encyclopedic Knowledge
Samer Hassan and Rada Mihalcea 1192

Joint Optimization for Machine Translation System Combination
Xiaodong He and Kristina Toutanova 1202

Fully Lexicalising CCGbank with Hat Categories
Matthew Honnibal and James R. Curran 1212

Bilingually-Constrained (Monolingual) Shift-Reduce Parsing
Liang Huang, Wenbin Jiang and Qun Liu 1222

Accurate Semantic Class Classifier for Coreference Resolution
Zhiheng Huang, Guangping Zeng, Weiqun Xu and Asli Celikyilmaz 1232

Real-Word Spelling Correction using Google Web 1T 3-grams
Aminul Islam and Diana Inkpen 1241

Semi-supervised Speech Act Recognition in Emails and Forums
Minwoo Jeong, Chin-Yew Lin and Gary Geunbae Lee 1250

Using Morphological and Syntactic Structures for Chinese Opinion Analysis
Lun-Wei Ku, Ting-Hao Huang and Hsin-Hsi Chen 1260

Finding Short Definitions of Terms on Web Pages
Gerasimos Lampouras and Ion Androutsopoulos 1270

Improving Nominal SRL in Chinese Language with Verbal SRL Information and Automatic Predicate Recognition
Junhui Li, Guodong Zhou, Hai Zhao, Qiaoming Zhu and Peide Qian 1280

On the Use of Virtual Evidence in Conditional Random Fields
Xiao Li 1289

Refining Grammars for Parsing with Hierarchical Semantic Knowledge
Xiaojun Lin, Yang Fan, Meng Zhang, Xihong Wu and Huisheng Chi 1298

Bayesian Learning of Phrasal Tree-to-String Templates
Ding Liu and Daniel Gildea 1308

Human-competitive Tagging Using Automatic Keyphrase Extraction
Olena Medelyan, Eibe Frank and Ian H. Witten 1318

Supervised Learning of a Probabilistic Lexicon of Verb Semantic Classes
Yusuke Miyao and Jun'ichi Tsujii 1328

A Study on the Semantic Relatedness of Query and Document Terms in Information Retrieval
Christof Müller and Iryna Gurevych 1338

Predicting Subjectivity in Multimodal Conversations
Gabriel Murray and Giuseppe Carenini 1348

Improved Statistical Machine Translation for Resource-Poor Languages Using Related Resource-Rich Languages
Preslav Nakov and Hwee Tou Ng 1358

What's in a Name? In Some Languages, Grammatical Gender
Vivi Nastase and Marius Popescu 1368

Convolution Kernels on Constituent, Dependency and Sequential Structures for Relation Extraction
Truc-Vien T. Nguyen, Alessandro Moschitti and Giuseppe Riccardi 1378

Automatic Acquisition of the Argument-Predicate *Relations from a Frame-Annotated Corpus*
Ekaterina Ovchinnikova, Theodore Alexandrov and Tonio Wandmacher 1388

Detecting Speculations and their Scopes in Scientific Text
Arzucan Özgür and Dragomir R. Radev 1398

Cross-Cultural Analysis of Blogs and Forums with Mixed-Collection Topic Models
Michael Paul and Roxana Girju 1408

Consensus Training for Consensus Decoding in Machine Translation
Adam Pauls, John Denero and Dan Klein 1418

Using Word-Sense Disambiguation Methods to Classify Web Queries by Intent
Emily Pitler and Ken Church 1428

Semi-Supervised Learning for Semantic Relation Classification using Stratified Sampling Strategy
Longhua Qian, Guodong Zhou, Fang Kong and Qiaoming Zhu 1437

Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis
Changqin Quan and Fuji Ren 1446

A Probabilistic Model for Associative Anaphora Resolution
Ryohei Sasano and Sadao Kurohashi 1455

Quantifier Scope Disambiguation Using Extracted Pragmatic Knowledge: Preliminary Results
Prakash Srinivasan and Alexander Yates 1465

Chinese Semantic Role Labeling with Shallow Parsing
Weiwei Sun, Zhifang Sui, Meng Wang and Xin Wang 1475

Discovery of Term Variation in Japanese Web Search Queries
Hisami Suzuki, Xiao Li and Jianfeng Gao 1484

Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics
Simone Teufel, Advaith Siddharthan and Colin Batchelor 1493

Character-level Analysis of Semi-Structured Documents for Set Expansion
Richard C. Wang and William W. Cohen 1503

Classifying Relations for Biomedical Named Entity Disambiguation
Xinglong Wang, Jun'ichi Tsujii and Sophia Ananiadou 1513

Domain Adaptive Bootstrapping for Named Entity Recognition
Dan Wu, Wee Sun Lee, Nan Ye and Hai Leong Chieu 1523

Phrase Dependency Parsing for Opinion Mining
Yuanbin Wu, Qi Zhang, Xuangjing Huang and Lide Wu 1533

Polynomial to Linear: Efficient Classification with Conjunctive Features
Naoki Yoshinaga and Masaru Kitsuregawa 1542

K-Best Combination of Syntactic Parsers
Hui Zhang, Min Zhang, Chew Lim Tan and Haizhou Li 1552

Chinese Novelty Mining
Yi Zhang and Flora S. Tsai 1561

Latent Document Re-Ranking
Dong Zhou and Vincent Wade 1571

Conference Program

Thursday, August 06, 2009

8:45–9:00 Opening remarks

9:00–10:00 Invited Talk

10:00–10:30 Coffee Break

Session 1A (Theatre): Semantic Parsing

10:30–10:55 *Unsupervised Semantic Parsing*
Hoifung Poon and Pedro Domingos

10:55–11:20 *Graph Alignment for Semi-Supervised Semantic Role Labeling*
Hagen Fürstenau and Mirella Lapata

11:20–11:45 *Semi-supervised Semantic Role Labeling Using the Latent Words Language Model*
Koen Deschacht and Marie-Francine Moens

11:45–12:10 *Semantic Dependency Parsing of NomBank and PropBank: An Efficient Integrated Approach via a Large-scale Feature Selection*
Hai Zhao, Wenliang Chen and Chunyu Kit

Session 1B (MR208): Machine Translation I

10:30–10:55 *First- and Second-Order Expectation Semirings with Applications to Minimum-Risk Training on Translation Forests*
Zhifei Li and Jason Eisner

10:55–11:20 *Feasibility of Human-in-the-loop Minimum Error Rate Training*
Omar F. Zaidan and Chris Callison-Burch

11:20–11:45 *Cube Pruning as Heuristic Search*
Mark Hopkins and Greg Langmead

11:45–12:10 *Effective Use of Linguistic and Contextual Information for Statistical Machine Translation*
Libin Shen, Jinxi Xu, Bing Zhang, Spyros Matsoukas and Ralph Weischedel

Thursday, August 06, 2009 (continued)

Session 1C (MR209): Machine Learning and Statistical Models I

10:30–10:55 *Active Learning by Labeling Features*
Gregory Druck, Burr Settles and Andrew McCallum

10:55–11:20 *Efficient Kernels for Sentence Pair Classification*
Fabio Massimo Zanzotto and Lorenzo Dell'Arciprete

11:20–11:45 *Graphical Models over Multiple Strings*
Markus Dreyer and Jason Eisner

11:45–12:10 *Reverse Engineering of Tree Kernel Feature Spaces*
Daniele Pighin and Alessandro Moschitti

Session 1D (MR203): Information Extraction

10:30–10:55 *A Rich Feature Vector for Protein-Protein Interaction Extraction from Multiple Corpora*
Makoto Miwa, Rune Sætre, Yusuke Miyao and Jun'ichi Tsujii

10:55–11:20 *Generalized Expectation Criteria for Bootstrapping Extractors using Record-Text Alignment*
Kedar Bellare and Andrew McCallum

11:20–11:45 *Nested Named Entity Recognition*
Jenny Rose Finkel and Christopher D. Manning

11:45–12:10 *A Unified Model of Phrasal and Sentential Evidence for Information Extraction*
Siddharth Patwardhan and Ellen Riloff

12:10–13:50 Lunch

Thursday, August 06, 2009 (continued)

Session 2A (Theatre): Subjectivity and Sentiment I

13:50–14:15 *Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm*
Jingjing Liu and Stephanie Seneff

14:15–14:40 *Supervised and Unsupervised Methods in Employing Discourse Relations for Improving Opinion Polarity Classification*
Swapna Somasundaran, Galileo Namata, Janyce Wiebe and Lise Getoor

14:40–15:05 *Sentiment Analysis of Conditional Sentences*
Ramanathan Narayanan, Bing Liu and Alok Choudhary

15:05–15:30 *Subjectivity Word Sense Disambiguation*
Cem Akkaya, Janyce Wiebe and Rada Mihalcea

Session 2B (MR208): Machine Translation II

13:50–14:15 *Non-Projective Parsing for Statistical Machine Translation*
Xavier Carreras and Michael Collins

14:15–14:40 *Extending Statistical Machine Translation with Discriminative and Trigger-Based Lexicon Models*
Arne Mauser, Saša Hasan and Hermann Ney

14:40–15:05 *Feature-Rich Translation by Quasi-Synchronous Lattice Parsing*
Kevin Gimpel and Noah A. Smith

15:05–15:30 *Improved Word Alignment with Statistics and Linguistic Heuristics*
Ulf Hermjakob

Thursday, August 06, 2009 (continued)

Session 2C (MR209): Natural Language Processing for Web 2.0

13:50–14:15 *Entity Extraction via Ensemble Semantics*
Marco Pennacchiotti and Patrick Pantel

14:15–14:40 *Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora*
Daniel Ramage, David Hall, Ramesh Nallapati and Christopher D. Manning

14:40–15:05 *Clustering to Find Exemplar Terms for Keyphrase Extraction*
Zhiyuan Liu, Peng Li, Yabin Zheng and Maosong Sun

15:05–15:30 *Geo-mining: Discovery of Road and Transport Networks Using Directional Patterns*
Dmitry Davidov and Ari Rappoport

Session 2D (MR203): Language Resources and Evaluation

13:50–14:15 *Wikipedia as Frame Information Repository*
Sara Tonelli and Claudio Giuliano

14:15–14:40 *Fast, Cheap, and Creative: Evaluating Translation Quality Using Amazon's Mechanical Turk*
Chris Callison-Burch

14:40–15:05 *How Well Does Active Learning* Actually *Work? Time-based Evaluation of Cost-reduction Strategies for Language Documentation*
Jason Baldridge and Alexis Palmer

15:05–15:30 *Automatically Evaluating Content Selection in Summarization without Human Models*
Annie Louis and Ani Nenkova

15:30–16:00 Coffee Break

Thursday, August 06, 2009 (continued)

Session 3A (Theatre): Discourse and Dialogue

16:00–16:25 *Classifier Combination for Contextual Idiom Detection Without Labelled Data*
Linlin Li and Caroline Sporleder

16:25–16:50 *Deriving Lexical and Syntactic Expectation-based Measures for Psycholinguistic Modeling via Incremental Top-down Parsing*
Brian Roark, Asaf Bachrach, Carlos Cardenas and Christophe Pallier

16:50–17:15 *It's Not You, it's Me: Detecting Flirting and its Misperception in Speed-Dates*
Rajesh Ranganath, Dan Jurafsky and Dan McFarland

17:15–17:40 *Recognizing Implicit Discourse Relations in the Penn Discourse Treebank*
Ziheng Lin, Min-Yen Kan and Hwee Tou Ng

Session 3B (MR208): Machine Translation III

16:00–16:25 *A Bayesian Model of Syntax-Directed Tree to String Grammar Induction*
Trevor Cohn and Phil Blunsom

16:25–16:50 *Better Synchronous Binarization for Machine Translation*
Tong Xiao, Mu Li, Dongdong Zhang, Jingbo Zhu and Ming Zhou

16:50–17:15 *Accuracy-Based Scoring for DOT: Towards Direct Error Minimization for Data-Oriented Translation*
Daniel Galron, Sergio Penkale, Andy Way and I. Dan Melamed

17:15–17:40 *Improved Statistical Machine Translation Using Monolingually-Derived Paraphrases*
Yuval Marton, Chris Callison-Burch and Philip Resnik

Thursday, August 06, 2009 (continued)

Session 3C (MR209): Summarization and Generation

16:00–16:25 *A Comparison of Model Free versus Model Intensive Approaches to Sentence Compression*
Tadashi Nomoto

16:25–16:50 *Natural Language Generation with Tree Conditional Random Fields*
Wei Lu, Hwee Tou Ng and Wee Sun Lee

16:50–17:15 *Perceptron Reranking for CCG Realization*
Michael White and Rajakrishnan Rajkumar

17:15–17:40 *Multi-Document Summarisation Using Generic Relation Extraction*
Ben Hachey

Session 3D (MR203): Lexical Semantics I

16:00–16:25 *Language Models Based on Semantic Composition*
Jeff Mitchell and Mirella Lapata

16:25–16:50 *Graded Word Sense Assignment*
Katrin Erk and Diana McCarthy

16:50–17:15 *Joint Learning of Preposition Senses and Semantic Roles of Prepositional Phrases*
Daniel Dahlmeier, Hwee Tou Ng and Tanja Schultz

17:15–17:40 *Projecting Parameters for Multilingual Word Sense Disambiguation*
Mitesh M. Khapra, Sapan Shah, Piyush Kedia and Pushpak Bhattacharyya

Thursday, August 06, 2009 (continued)

18:00–20:00 **Poster Session and Reception**

Gazpacho and Summer Rash: Lexical Relationships from Temporal Patterns of Web Search Queries
Enrique Alfonseca, Massimiliano Ciaramita and Keith Hall

A Compact Forest for Scalable Inference over Entailment and Paraphrase Rules
Roy Bar-Haim, Jonathan Berant and Ido Dagan

Discriminative Substring Decoding for Transliteration
Colin Cherry and Hisami Suzuki

Re-Ranking Models Based-on Small Training Data for Spoken Language Understanding
Marco Dinarelli, Alessandro Moschitti and Giuseppe Riccardi

Empirical Exploitation of Click Data for Task Specific Ranking
Anlei Dong, Yi Chang, Shihao Ji, Ciya Liao, Xin Li and Zhaohui Zheng

The Feature Subspace Method for SMT System Combination
Nan Duan, Mu Li, Tong Xiao and Ming Zhou

Lattice-based System Combination for Statistical Machine Translation
Yang Feng, Yang Liu, Haitao Mi, Qun Liu and Yajuan Lü

A Joint Language Model With Fine-grain Syntactic Tags
Denis Filimonov and Mary Harper

Bidirectional Phrase-based Statistical Machine Translation
Andrew Finch and Eiichiro Sumita

Real-time Decision Detection in Multi-party Dialogue
Matthew Frampton, Jia Huang, Trung Bui and Stanley Peters

On the Role of Lexical Features in Sequence Labeling
Yoav Goldberg and Michael Elhadad

Thursday, August 06, 2009 (continued)

18:00–20:00 **Poster Session and Reception (continued)**

Simple Coreference Resolution with Rich Syntactic and Semantic Features
Aria Haghighi and Dan Klein

Descriptive and Empirical Approaches to Capturing Underlying Dependencies among Parsing Errors
Tadayoshi Hara, Yusuke Miyao and Jun'ichi Tsujii

Large-Scale Verb Entailment Acquisition from the Web
Chikara Hashimoto, Kentaro Torisawa, Kow Kuroda, Stijn De Saeger, Masaki Murata and Jun'ichi Kazama

A Syntactified Direct Translation Model with Linear-time Decoding
Hany Hassan, Khalil Sima'an and Andy Way

Cross-lingual Semantic Relatedness Using Encyclopedic Knowledge
Samer Hassan and Rada Mihalcea

Joint Optimization for Machine Translation System Combination
Xiaodong He and Kristina Toutanova

Fully Lexicalising CCGbank with Hat Categories
Matthew Honnibal and James R. Curran

Bilingually-Constrained (Monolingual) Shift-Reduce Parsing
Liang Huang, Wenbin Jiang and Qun Liu

Accurate Semantic Class Classifier for Coreference Resolution
Zhiheng Huang, Guangping Zeng, Weiqun Xu and Asli Celikyilmaz

Real-Word Spelling Correction using Google Web 1T 3-grams
Aminul Islam and Diana Inkpen

Semi-supervised Speech Act Recognition in Emails and Forums
Minwoo Jeong, Chin-Yew Lin and Gary Geunbae Lee

Thursday, August 06, 2009 (continued)

18:00–20:00 **Poster Session and Reception (continued)**

Using Morphological and Syntactic Structures for Chinese Opinion Analysis
Lun-Wei Ku, Ting-Hao Huang and Hsin-Hsi Chen

Finding Short Definitions of Terms on Web Pages
Gerasimos Lampouras and Ion Androutsopoulos

Improving Nominal SRL in Chinese Language with Verbal SRL Information and Automatic Predicate Recognition
Junhui Li, Guodong Zhou, Hai Zhao, Qiaoming Zhu and Peide Qian

On the Use of Virtual Evidence in Conditional Random Fields
Xiao Li

Refining Grammars for Parsing with Hierarchical Semantic Knowledge
Xiaojun Lin, Yang Fan, Meng Zhang, Xihong Wu and Huisheng Chi

Bayesian Learning of Phrasal Tree-to-String Templates
Ding Liu and Daniel Gildea

Human-competitive Tagging Using Automatic Keyphrase Extraction
Olena Medelyan, Eibe Frank and Ian H. Witten

Supervised Learning of a Probabilistic Lexicon of Verb Semantic Classes
Yusuke Miyao and Jun'ichi Tsujii

A Study on the Semantic Relatedness of Query and Document Terms in Information Retrieval
Christof Müller and Iryna Gurevych

Predicting Subjectivity in Multimodal Conversations
Gabriel Murray and Giuseppe Carenini

Improved Statistical Machine Translation for Resource-Poor Languages Using Related Resource-Rich Languages
Preslav Nakov and Hwee Tou Ng

Thursday, August 06, 2009 (continued)

18:00–20:00 **Poster Session and Reception (continued)**

What's in a Name? In Some Languages, Grammatical Gender
Vivi Nastase and Marius Popescu

Convolution Kernels on Constituent, Dependency and Sequential Structures for Relation Extraction
Truc-Vien T. Nguyen, Alessandro Moschitti and Giuseppe Riccardi

Automatic Acquisition of the Argument-Predicate *Relations from a Frame-Annotated Corpus*
Ekaterina Ovchinnikova, Theodore Alexandrov and Tonio Wandmacher

Detecting Speculations and their Scopes in Scientific Text
Arzucan Özgür and Dragomir R. Radev

Cross-Cultural Analysis of Blogs and Forums with Mixed-Collection Topic Models
Michael Paul and Roxana Girju

Consensus Training for Consensus Decoding in Machine Translation
Adam Pauls, John Denero and Dan Klein

Using Word-Sense Disambiguation Methods to Classify Web Queries by Intent
Emily Pitler and Ken Church

Semi-Supervised Learning for Semantic Relation Classification using Stratified Sampling Strategy
Longhua Qian, Guodong Zhou, Fang Kong and Qiaoming Zhu

Construction of a Blog Emotion Corpus for Chinese Emotional Expression Analysis
Changqin Quan and Fuji Ren

A Probabilistic Model for Associative Anaphora Resolution
Ryohei Sasano and Sadao Kurohashi

Quantifier Scope Disambiguation Using Extracted Pragmatic Knowledge: Preliminary Results
Prakash Srinivasan and Alexander Yates

Thursday, August 06, 2009 (continued)

18:00–20:00 **Poster Session and Reception (continued)**

Chinese Semantic Role Labeling with Shallow Parsing
Weiwei Sun, Zhifang Sui, Meng Wang and Xin Wang

Discovery of Term Variation in Japanese Web Search Queries
Hisami Suzuki, Xiao Li and Jianfeng Gao

Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics
Simone Teufel, Advaith Siddharthan and Colin Batchelor

Character-level Analysis of Semi-Structured Documents for Set Expansion
Richard C. Wang and William W. Cohen

Classifying Relations for Biomedical Named Entity Disambiguation
Xinglong Wang, Jun'ichi Tsujii and Sophia Ananiadou

Domain adaptive bootstrapping for named entity recognition
Dan Wu, Wee Sun Lee, Nan Ye and Hai Leong Chieu

Phrase Dependency Parsing for Opinion Mining
Yuanbin Wu, Qi Zhang, Xuangjing Huang and Lide Wu

Polynomial to Linear: Efficient Classification with Conjunctive Features
Naoki Yoshinaga and Masaru Kitsuregawa

K-Best Combination of Syntactic Parsers
Hui Zhang, Min Zhang, Chew Lim Tan and Haizhou Li

Chinese Novelty Mining
Yi Zhang and Flora S. Tsai

Latent Document Re-Ranking
Dong Zhou and Vincent Wade

Friday, August 07, 2009

Session 4A (Theatre): Multi-word Expressions

8:45–9:10 *Multi-Word Expression Identification Using Sentence Surface Features*
Ram Boukobza and Ari Rappoport

9:10–9:35 *Acquiring Translation Equivalences of Multiword Expressions by Normalized Correlation Frequencies*
Ming-Hong Bai, Jia-Ming You, Keh-Jiann Chen and Jason S. Chang

9:35–10:00 *Collocation Extraction Using Monolingual Word Alignment Method*
Zhanyi Liu, Haifeng Wang, Hua Wu and Sheng Li

Session 4B (MR208): Machine Learning and Statistical Models II

8:45–9:10 *Multi-Class Confidence Weighted Algorithms*
Koby Crammer, Mark Dredze and Alex Kulesza

9:10–9:35 *Model Adaptation via Model Interpolation and Boosting for Web Search Ranking*
Jianfeng Gao, Qiang Wu, Chris Burges, Krysta Svore, Yi Su, Nazan Khan, Shalin Shah and Hongyan Zhou

9:35–10:00 *A Structural Support Vector Method for Extracting Contexts and Answers of Questions from Online Forums*
Wen-Yun Yang, Yunbo Cao and Chin-Yew Lin

Session 4C (MR209): Information Retrieval and Questions Answering

8:45–9:10 *Mining Search Engine Clickthrough Log for Matching N-gram Features*
Huihsin Tseng, Longbin Chen, Fan Li, Ziming Zhuang, Lei Duan and Belle Tseng

9:10–9:35 *The Role of Named Entities in Web People Search*
Javier Artiles, Enrique Amigó and Julio Gonzalo

9:35–10:00 *Investigation of Question Classifier in Question Answering*
Zhiheng Huang, Marcus Thint and Asli Celikyilmaz

Friday, August 07, 2009 (continued)

Session 4D (MR203): Syntax and Parsing I

8:45–9:10 *An Empirical Study of Semi-supervised Structured Conditional Models for Dependency Parsing*
Jun Suzuki, Hideki Isozaki, Xavier Carreras and Michael Collins

9:10–9:35 *Statistical Bistratal Dependency Parsing*
Richard Johansson

9:35–10:00 *Improving Dependency Parsing with Subtrees from Auto-Parsed Data*
Wenliang Chen, Jun'ichi Kazama, Kiyotaka Uchimoto and Kentaro Torisawa

10:00–10:30 Coffee Break

Session 5A (Theatre): Subjectivity and Sentiment II

10:30–10:55 *Topic-wise, Sentiment-wise, or Otherwise? Identifying the Hidden Dimension for Unsupervised Text Classification*
Sajib Dasgupta and Vincent Ng

10:55–11:20 *Adapting a Polarity Lexicon using Integer Linear Programming for Domain-Specific Sentiment Classification*
Yejin Choi and Claire Cardie

11:20–11:45 *Generating High-Coverage Semantic Orientation Lexicons From Overtly Marked Words and a Thesaurus*
Saif Mohammad, Cody Dunne and Bonnie Dorr

11:45–12:10 *Matching Reviews to Objects using a Language Model*
Nilesh Dalvi, Ravi Kumar, Bo Pang and Andrew Tomkins

Friday, August 07, 2009 (continued)

Session 5B (MR208): Lexical Semantics II

10:30–10:55 *EEG Responds to Conceptual Stimuli and Corpus Semantics*
Brian Murphy, Marco Baroni and Massimo Poesio

10:55–11:20 *A Comparison of Windowless and Window-Based Computational Association Measures as Predictors of Syntagmatic Human Associations*
Justin Washtell and Katja Markert

11:20–11:45 *Improving Verb Clustering with Automatically Acquired Selectional Preferences*
Lin Sun and Anna Korhonen

11:45–12:10 *Improving Web Search Relevance with Semantic Features*
Yumao Lu, Fuchun Peng, Gilad Mishne, Xing Wei and Benoit Dumoulin

Session 5C (MR209): Phonology and Morphology

10:30–10:55 *Can Chinese Phonemes Improve Machine Transliteration?: A Comparative Study of English-to-Chinese Transliteration Models*
Jong-Hoon Oh, Kiyotaka Uchimoto and Kentaro Torisawa

10:55–11:20 *Unsupervised Morphological Segmentation and Clustering with Document Boundaries*
Taesun Moon, Katrin Erk and Jason Baldridge

11:20–11:45 *The Infinite HMM for Unsupervised PoS Tagging*
Jurgen Van Gael, Andreas Vlachos and Zoubin Ghahramani

11:45–12:10 *A Simple Unsupervised Learner for POS Disambiguation Rules Given Only a Minimal Lexicon*
Qiuye Zhao and Mitch Marcus

Friday, August 07, 2009 (continued)

Session 5D (MR203): Machine Translation IV

10:30–10:55 *Tree Kernel-based SVM with Structured Syntactic Knowledge for BTG-based Phrase Reordering*
Min Zhang and Haizhou Li

10:55–11:20 *Discriminative Corpus Weight Estimation for Machine Translation*
Spyros Matsoukas, Antti-Veikko I. Rosti and Bing Zhang

11:20–11:45 *Unsupervised Tokenization for Machine Translation*
Tagyoung Chung and Daniel Gildea

11:45–12:10 *Synchronous Tree Adjoining Machine Translation*
Steve DeNeefe and Kevin Knight

12:10–13:50 Lunch

Session 6A (Theatre): Speech and Language Modeling

13:50–14:15 *Word Buffering Models for Improved Speech Repair Parsing*
Tim Miller

14:15–14:40 *Less is More: Significance-Based N-gram Selection for Smaller, Better Language Models*
Robert C. Moore and Chris Quirk

14:40–15:05 *Stream-based Randomised Language Models for SMT*
Abby Levenberg and Miles Osborne

15:05–15:30 *Integrating Sentence- and Word-level Error Identification for Disfluency Correction*
Erin Fitzgerald, Frederick Jelinek and Keith Hall

Friday, August 07, 2009 (continued)

Session 6B (MR208): Semantic Similarity

13:50–14:15 *Estimating Semantic Distance Using Soft Semantic Constraints in Knowledge-Source – Corpus Hybrid Models*
Yuval Marton, Saif Mohammad and Philip Resnik

14:15–14:40 *Recognizing Textual Relatedness with Predicate-Argument Structures*
Rui Wang and Yi Zhang

14:40–15:05 *Learning Term-weighting Functions for Similarity Measures*
Wen-tau Yih

15:05–15:30 *A Relational Model of Semantic Similarity between Words using Automatically Extracted Lexical Pattern Clusters from the Web*
Danushka Bollegala, Yutaka Matsuo and Mitsuru Ishizuka

Session 6C (MR209): Syntax and Parsing II

13:50–14:15 *Unbounded Dependency Recovery for Parser Evaluation*
Laura Rimell, Stephen Clark and Mark Steedman

14:15–14:40 *Parser Adaptation and Projection with Quasi-Synchronous Grammar Features*
David A. Smith and Jason Eisner

14:40–15:05 *Self-Training PCFG Grammars with Latent Annotations Across Languages*
Zhongqiang Huang and Mary Harper

15:05–15:30 *An Alternative to Head-Driven Approaches for Parsing a (Relatively) Free Word-Order Language*
Reut Tsarfaty, Khalil Sima'an and Remko Scha

Friday, August 07, 2009 (continued)

Session 6D (MR203): Multilinguality

13:50–14:15 *Enhancement of Lexical Concepts Using Cross-lingual Web Mining*
Dmitry Davidov and Ari Rappoport

14:15–14:40 *Bilingual Dictionary Generation for Low-resourced Language Pairs*
István Varga and Shoichi Yokoyama

14:40–15:05 *Multilingual Spectral Clustering Using Document Similarity Propagation*
Dani Yogatama and Kumiko Tanaka-Ishii

15:05–15:30 *Polylingual Topic Models*
David Mimno, Hanna M. Wallach, Jason Naradowsky, David A. Smith and Andrew McCallum

15:30–16:00 Coffee Break

Session 7A (Theatre): Natural Language Applications

16:00–16:25 *Using the Web for Language Independent Spellchecking and Autocorrection*
Casey Whitelaw, Ben Hutchinson, Grace Y Chung and Ged Ellis

16:25–16:50 *Statistical Estimation of Word Acquisition with Application to Readability Prediction*
Paul Kidwell, Guy Lebanon and Kevyn Collins-Thompson

16:50–17:15 *Combining Collocations, Lexical and Encyclopedic Knowledge for Metonymy Resolution*
Vivi Nastase and Michael Strube

17:15–17:40 *Segmenting Email Message Text into Zones*
Andrew Lampert, Robert Dale and Cécile Paris

Friday, August 07, 2009 (continued)

Session 7B (MR208): Lexical Semantics III

16:00–16:25 *Hypernym Discovery Based on Distributional Similarity and Hierarchical Structures*
Ichiro Yamada, Kentaro Torisawa, Jun'ichi Kazama, Kow Kuroda, Masaki Murata, Stijn De Saeger, Francis Bond and Asuka Sumida

16:25–16:50 *Web-Scale Distributional Similarity and Entity Set Expansion*
Patrick Pantel, Eric Crestan, Arkady Borkovsky, Ana-Maria Popescu and Vishnu Vyas

16:50–17:15 *Toward Completeness in Concept Extraction and Classification*
Eduard Hovy, Zornitsa Kozareva and Ellen Riloff

17:15–17:40 *Reading to Learn: Constructing Features from Semantic Abstracts*
Jacob Eisenstein, James Clarke, Dan Goldwasser and Dan Roth

Session 7C (MR209): Coreference Resolution

16:00–16:25 *Supervised Models for Coreference Resolution*
Altaf Rahman and Vincent Ng

16:25–16:50 *Global Learning of Noun Phrase Anaphoricity in Coreference Resolution via Label Propagation*
GuoDong Zhou and Fang Kong

16:50–17:15 *Employing the Centering Theory in Pronoun Resolution from the Semantic Perspective*
Fang Kong, GuoDong Zhou and Qiaoming Zhu

17:15–17:40 *Person Cross Document Coreference with Name Perplexity Estimates*
Octavian Popescu

Friday, August 07, 2009 (continued)

Session 7D (MR203): Machine Translation V

16:00–16:25 *Learning Linear Ordering Problems for Better Translation*
Roy Tromble and Jason Eisner

16:25–16:50 *Weighted Alignment Matrices for Statistical Machine Translation*
Yang Liu, Tian Xia, Xinyan Xiao and Qun Liu

16:50–17:15 *Sinuhe – Statistical Machine Translation using a Globally Trained Conditional Exponential Family Translation Model*
Matti Kääriäinen

17:15–17:40 *Fast Translation Rule Matching for Syntax-based Statistical Machine Translation*
Hui Zhang, Min Zhang, Haizhou Li and Chew Lim Tan