ACL-08: HLT

46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies

Proceedings of the Conference

June 15–20, 2008
The Ohio State University
Columbus, Ohio, USA

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

©2008 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 978-1-932432-04-6

Table of Contents

Preface: General Chair xi

Preface: Program Chairs xii

Organizers xv

Program Committee xvii

Conference Program xxi

Mining Wiki Resources for Multilingual Named Entity Recognition
Alexander E. Richman and Patrick Schone 1

Distributional Identification of Non-Referential Pronouns
Shane Bergsma, Dekang Lin and Randy Goebel 10

Weakly-Supervised Acquisition of Open-Domain Classes and Class Attributes from Web Documents and Query Logs
Marius Paşca and Benjamin Van Durme 19

The Tradeoffs Between Open and Traditional Relation Extraction
Michele Banko and Oren Etzioni 28

PDT 2.0 Requirements on a Query Language
Jiří Mírovský 37

Task-oriented Evaluation of Syntactic Parsers and Their Representations
Yusuke Miyao, Rune Sætre, Kenji Sagae, Takuya Matsuzaki and Jun'ichi Tsujii 46

MAXSIM: A Maximum Similarity Metric for Machine Translation Evaluation
Yee Seng Chan and Hwee Tou Ng 55

Contradictions and Justifications: Extensions to the Textual Entailment Task
Ellen M. Voorhees 63

Cohesive Phrase-Based Decoding for Statistical Machine Translation
Colin Cherry 72

Phrase Table Training for Precision and Recall: What Makes a Good Phrase and a Good Phrase Pair?
Yonggang Deng, Jia Xu and Yuqing Gao 81

Measure Word Generation for English-Chinese SMT Systems
Dongdong Zhang, Mu Li, Nan Duan, Chi-Ho Li and Ming Zhou 89

Bayesian Learning of Non-Compositional Phrases with Synchronous Parsing
Hao Zhang, Chris Quirk, Robert C. Moore and Daniel Gildea 97

Applying a Grammar-Based Language Model to a Simplified Broadcast-News Transcription Task
Tobias Kaufmann and Beat Pfister . . . 106

Automatic Editing in a Back-End Speech-to-Text System
Maximilian Bisani, Paul Vozila, Olivier Divay and Jeff Adams . . . 114

Grounded Language Modeling for Automatic Speech Recognition of Sports Video
Michael Fleischman and Deb Roy . . . 121

Lexicalized Phonotactic Word Segmentation
Margaret M. Fleck . . . 130

A Re-examination of Query Expansion Using Lexical Resources
Hui Fang . . . 139

Selecting Query Term Alternations for Web Search by Exploiting Query Contexts
Guihong Cao, Stephen Robertson and Jian-Yun Nie . . . 148

Searching Questions by Identifying Question Topic and Question Focus
Huizhong Duan, Yunbo Cao, Chin-Yew Lin and Yong Yu . . . 156

Trainable Generation of Big-Five Personality Styles through Data-Driven Parameter Estimation
François Mairesse and Marilyn Walker . . . 165

Correcting Misuse of Verb Forms
John Lee and Stephanie Seneff . . . 174

Hypertagging: Supertagging for Surface Realization with CCG
Dominic Espinosa, Michael White and Dennis Mehay . . . 183

Forest-Based Translation
Haitao Mi, Liang Huang and Qun Liu . . . 192

A Discriminative Latent Variable Model for Statistical Machine Translation
Phil Blunsom, Trevor Cohn and Miles Osborne . . . 200

Efficient Multi-Pass Decoding for Synchronous Context Free Grammars
Hao Zhang and Daniel Gildea . . . 209

Regular Tree Grammars as a Formalism for Scope Underspecification
Alexander Koller, Michaela Regneri and Stefan Thater . . . 218

Classification of Semantic Relationships between Nominals Using Pattern Clusters
Dmitry Davidov and Ari Rappoport . . . 227

Vector-based Models of Semantic Composition
Jeff Mitchell and Mirella Lapata . . . 236

Exploiting Feature Hierarchy for Transfer Learning in Named Entity Recognition
Andrew Arnold, Ramesh Nallapati and William W. Cohen . . . 245

Refining Event Extraction through Cross-Document Inference
Heng Ji and Ralph Grishman 254

Learning Document-Level Semantic Properties from Free-Text Annotations
S.R.K. Branavan, Harr Chen, Jacob Eisenstein and Regina Barzilay 263

Automatic Image Annotation Using Auxiliary Text Information
Yansong Feng and Mirella Lapata 272

Hedge Classification in Biomedical Texts with a Weakly Supervised Selection of Keywords
György Szarvas 281

When Specialists and Generalists Work Together: Overcoming Domain Dependence in Sentiment Tagging
Alina Andreevskaia and Sabine Bergler 290

A Generic Sentence Trimmer with CRFs
Tadashi Nomoto 299

A Joint Model of Text and Aspect Ratings for Sentiment Summarization
Ivan Titov and Ryan McDonald 308

Improving Parsing and PP Attachment Performance with Sense Information
Eneko Agirre, Timothy Baldwin and David Martinez 317

A Logical Basis for the D Combinator and Normal Form in CCG
Frederick Hoyt and Jason Baldridge 326

Parsing Noun Phrase Structure with CCG
David Vadas and James R. Curran 335

Sentence Simplification for Semantic Role Labeling
David Vickrey and Daphne Koller 344

Summarizing Emails with Conversational Cohesion and Subjectivity
Giuseppe Carenini, Raymond T. Ng and Xiaodong Zhou 353

Ad Hoc Treebank Structures
Markus Dickinson 362

A Single Generative Model for Joint Morphological Segmentation and Syntactic Parsing
Yoav Goldberg and Reut Tsarfaty 371

Which Words Are Hard to Recognize? Prosodic, Lexical, and Disfluency Factors that Increase ASR Error Rates
Sharon Goldwater, Dan Jurafsky and Christopher D. Manning 380

Name Translation in Statistical Machine Translation - Learning When to Transliterate
Ulf Hermjakob, Kevin Knight and Hal Daumé III 389

Using Adaptor Grammars to Identify Synergies in the Unsupervised Acquisition of Linguistic Structure
Mark Johnson 398

Inducing Gazetteers for Named Entity Recognition by Large-Scale Clustering of Dependency Relations
Jun'ichi Kazama and Kentaro Torisawa 407

Evaluating Roget's Thesauri
Alistair Kennedy and Stan Szpakowicz 416

Unsupervised Translation Induction for Chinese Abbreviations using Monolingual Corpora
Zhifei Li and David Yarowsky 425

Which Are the Best Features for Automatic Verb Classification
Jianguo Li and Chris Brew 434

Collecting a Why-Question Corpus for Development and Evaluation of an Automatic QA-System
Joanna Mrozinski, Edward Whittaker and Sadaoki Furui 443

Solving Relational Similarity Problems Using the Web as a Corpus
Preslav Nakov and Marti A. Hearst 452

Combining Speech Retrieval Results with Generalized Additive Models
J. Scott Olsson and Douglas W. Oard 461

A Critical Reassessment of Evaluation Baselines for Speech Summarization
Gerald Penn and Xiaodan Zhu 470

Intensional Summaries as Cooperative Responses in Dialogue: Automation and Evaluation
Joseph Polifroni and Marilyn Walker 479

Word Clustering and Word Selection Based Feature Reduction for MaxEnt Based Hindi NER
Sujan Kumar Saha, Pabitra Mitra and Sudeshna Sarkar 488

Combining EM Training and the MDL Principle for an Automatic Verb Classification Incorporating Selectional Preferences
Sabine Schulte im Walde, Christian Hying, Christian Scheible and Helmut Schmid 496

Randomized Language Models via Perfect Hash Functions
David Talbot and Thorsten Brants 505

Applying Morphology Generation Models to Machine Translation
Kristina Toutanova, Hisami Suzuki and Achim Ruopp 514

Multilingual Harvesting of Cross-Cultural Stereotypes
Tony Veale, Yanfen Hao and Guofu Li 523

Semi-Supervised Convex Training for Dependency Parsing
Qin Iris Wang, Dale Schuurmans and Dekang Lin 532

Chinese-English Backward Transliteration Assisted with Mining Monolingual Web Pages
Fan Yang, Jun Zhao, Bo Zou, Kang Liu and Feifan Liu 541

Robustness and Generalization of Role Sets: PropBank vs. VerbNet
Beñat Zapirain, Eneko Agirre and Lluís Màrquez 550

A Tree Sequence Alignment-based Tree-to-Tree Translation Model
Min Zhang, Hongfei Jiang, Aiti Aw, Haizhou Li, Chew Lim Tan and Sheng Li 559

Automatic Syllabification with Structured SVMs for Letter-to-Phoneme Conversion
Susan Bartlett, Grzegorz Kondrak and Colin Cherry 568

A New String-to-Dependency Machine Translation Algorithm with a Target Dependency Language Model
Libin Shen, Jinxi Xu and Ralph Weischedel 577

Forest Reranking: Discriminative Parsing with Non-Local Features
Liang Huang 586

Simple Semi-supervised Dependency Parsing
Terry Koo, Xavier Carreras and Michael Collins 595

Optimal k-arization of Synchronous Tree-Adjoining Grammar
Rebecca Nesson, Giorgio Satta and Stuart M. Shieber 604

Enhancing Performance of Lexicalised Grammars
Rebecca Dridan, Valia Kordoni and Jeremy Nicholson 613

Assessing Dialog System User Simulation Evaluation Measures Using Human Judges
Hua Ai and Diane J. Litman 622

Robust Dialog Management with N-Best Hypotheses Using Dialog Examples and Agenda
Cheongjae Lee, Sangkeun Jung and Gary Geunbae Lee 630

Learning Effective Multimodal Dialogue Strategies from Wizard-of-Oz Data: Bootstrapping and Evaluation
Verena Rieser and Oliver Lemon 638

Phrase Chunking Using Entropy Guided Transformation Learning
Ruy Luiz Milidiú, Cícero Nogueira dos Santos and Julio C. Duarte 647

Learning Bigrams from Unigrams
Xiaojin Zhu, Andrew B. Goldberg, Michael Rabbat and Robert Nowak 656

Semi-Supervised Sequential Labeling and Segmentation Using Giga-Word Scale Unlabeled Data
Jun Suzuki and Hideki Isozaki 665

Large Scale Acquisition of Paraphrases for Learning Surface Patterns
Rahul Bhagat and Deepak Ravichandran 674

Contextual Preferences
Idan Szpektor, Ido Dagan, Roy Bar-Haim and Jacob Goldberger 683

Unsupervised Discovery of Generic Relationships Using Pattern Clusters and its Evaluation by Automatically Generated SAT Analogy Questions
Dmitry Davidov and Ari Rappoport 692

Improving Search Results Quality by Customizing Summary Lengths
Michael Kaisser, Marti A. Hearst and John B. Lowe 701

Using Conditional Random Fields to Extract Contexts and Answers of Questions from Online Forums
Shilin Ding, Gao Cong, Chin-Yew Lin and Xiaoyan Zhu 710

Learning to Rank Answers on Large Online QA Collections
Mihai Surdeanu, Massimiliano Ciaramita and Hugo Zaragoza 719

Unsupervised Lexicon-Based Resolution of Unknown Words for Full Morphological Analysis
Meni Adler, Yoav Goldberg, David Gabay and Michael Elhadad 728

Unsupervised Multilingual Learning for Morphological Segmentation
Benjamin Snyder and Regina Barzilay 737

EM Can Find Pretty Good HMM POS-Taggers (When Given a Good Start)
Yoav Goldberg, Meni Adler and Michael Elhadad 746

Distributed Word Clustering for Large Scale Class-Based Language Modeling in Machine Translation
Jakob Uszkoreit and Thorsten Brants 755

Enriching Morphologically Poor Languages for Statistical Machine Translation
Eleftherios Avramidis and Philipp Koehn 763

Learning Bilingual Lexicons from Monolingual Corpora
Aria Haghighi, Percy Liang, Taylor Berg-Kirkpatrick and Dan Klein 771

Pivot Approach for Extracting Paraphrase Patterns from Bilingual Corpora
Shiqi Zhao, Haifeng Wang, Ting Liu and Sheng Li 780

Unsupervised Learning of Narrative Event Chains
Nathanael Chambers and Dan Jurafsky 789

Semantic Role Labeling Systems for Arabic using Kernel Methods
Mona Diab, Alessandro Moschitti and Daniele Pighin 798

An Unsupervised Approach to Biography Production Using Wikipedia
Fadi Biadsy, Julia Hirschberg and Elena Filatova 807

Generating Impact-Based Summaries for Scientific Literature
Qiaozhu Mei and ChengXiang Zhai 816

Can You Summarize This? Identifying Correlates of Input Difficulty for Multi-Document Summarization
Ani Nenkova and Annie Louis 825

You Talking to Me? A Corpus and Algorithm for Conversation Disentanglement
Micha Elsner and Eugene Charniak 834

An Entity-Mention Model for Coreference Resolution with Inductive Logic Programming
Xiaofeng Yang, Jian Su, Jun Lang, Chew Lim Tan, Ting Liu and Sheng Li 843

Gestural Cohesion for Topic Segmentation
Jacob Eisenstein, Regina Barzilay and Randall Davis 852

Multi-Task Active Learning for Linguistic Annotations
Roi Reichart, Katrin Tomanek, Udo Hahn and Ari Rappoport 861

Generalized Expectation Criteria for Semi-Supervised Learning of Conditional Random Fields
Gideon S. Mann and Andrew McCallum 870

Analyzing the Errors of Unsupervised Learning
Percy Liang and Dan Klein 879

Joint Word Segmentation and POS Tagging Using a Single Perceptron
Yue Zhang and Stephen Clark 888

A Cascaded Linear Model for Joint Chinese Word Segmentation and Part-of-Speech Tagging
Wenbin Jiang, Liang Huang, Qun Liu and Yajuan Lü 897

Joint Processing and Discriminative Training for Letter-to-Phoneme Conversion
Sittichai Jiampojamarn, Colin Cherry and Grzegorz Kondrak 905

A Probabilistic Model for Fine-Grained Expert Search
Shenghua Bao, Huizhong Duan, Qi Zhou, Miao Xiong, Yunbo Cao and Yong Yu 914

Credibility Improves Topical Blog Post Retrieval
Wouter Weerkamp and Maarten de Rijke 923

Linguistically Motivated Features for Enhanced Back-of-the-Book Indexing
Andras Csomai and Rada Mihalcea 932

Resolving Personal Names in Email Using Context Expansion
Tamer Elsayed, Douglas W. Oard and Galileo Namata 941

Integrating Graph-Based and Transition-Based Dependency Parsers
Joakim Nivre and Ryan McDonald 950

Efficient, Feature-based, Conditional Random Field Parsing
Jenny Rose Finkel, Alex Kleeman and Christopher D. Manning 959

A Deductive Approach to Dependency Parsing
Carlos Gómez-Rodríguez, John Carroll and David Weir 968

Evaluating a Crosslinguistic Grammar Resource: A Case Study of Wambaya
Emily M. Bender 977

Better Alignments = Better Translations?
Kuzman Ganchev, João V. Graça and Ben Taskar 986

Mining Parenthetical Translations from the Web by Word Alignment
Dekang Lin, Shaojun Zhao, Benjamin Van Durme and Marius Paşca 994

Soft Syntactic Constraints for Hierarchical Phrased-Based Translation
Yuval Marton and Philip Resnik 1003

Generalizing Word Lattice Translation
Christopher Dyer, Smaranda Muresan and Philip Resnik 1012

Combining Multiple Resources to Improve SMT-based Paraphrasing Model
Shiqi Zhao, Cheng Niu, Ming Zhou, Ting Liu and Sheng Li 1021

Extraction of Entailed Semantic Relations Through Syntax-Based Comma Resolution
Vivek Srikumar, Roi Reichart, Mark Sammons, Ari Rappoport and Dan Roth 1030

Finding Contradictions in Text
Marie-Catherine de Marneffe, Anna N. Rafferty and Christopher D. Manning 1039

Semantic Class Learning from the Web with Hyponym Pattern Linkage Graphs
Zornitsa Kozareva, Ellen Riloff and Eduard Hovy 1048

Preface: General Chair

I am honored to be serving as General Conference Chair for the annual conference in our field. This year's conference, ACL-08: HLT, is jointly sponsored by the Association for Computational Linguistics and the North American Chapter of the Association for Computational Linguistics and it thus brings together the traditions of both organizations. As is evident from the title, one of those traditions is the focus on research from all areas of Human Language Technology, including information retrieval, natural language processing and speech. The conference features invited speakers in speech and information retrieval and there are sessions devoted to all three of these areas. I hope this conference will again encourage interaction among researchers from the different areas.

Since I was last involved in organizing the ACL Conferences back in the 90's, the conferences have grown dramatically. I was surprised to learn the number of people required to make the conference happen. Some 30 odd people are serving in Chair or Co-Chair capacity of various aspects of the conference. While I was pleased to have the opportunity of shaping aspects of the conference, I have to say that the real bulk of the work is done by the many Chairs involved. So I want to express my gratitude to all of them for their commitment and dedication to making sure that all ran smoothly. I am impressed by the energy and time that everyone gave to this volunteer activity.

I would like to thank the Program Chairs, Johanna Moore, Simone Teufel, James Allan and Sadaoki Furui, who have put in many hours to provide us with the main program for the conference and the Local Arrangements Chair, Chris Brew, who has provided us with the venue for the conference and oversaw the many time-demanding details. DJ Hovermale also put in many hours as webmaster, collecting information from everyone. I would like to thank the Chairs of the Student Research Workshop, Ebru Arisoy, Wolfgang Maier and Keisuke Inoue, who worked quite independently, along with the Faculty Advisor, Jan Wiebe. The Workshop Chair, Ming Zhou, managed the workshop program with ease, a program that has grown over the years so that it seems like a conference in and of itself. The Tutorial Chairs, Ani Nenkova, Marilyn Walker and Eugene Agichtein, have put together a fine tutorial program and the Demo Chair, Jimmy Lin, has organized a nice series of demos. The Sponsorship Chairs are responsible for bringing in funding to cover various programs and I would like to thank Inderjeet Mani, Josef van Genabith and Michael White for their efforts in this regard. The Publicity Chairs, Hal Daumé III, Eric Fosler-Lussier and Diane Kelly, reached out to communities outside the central natural language areas to encourage people to submit papers and attend the conference. Finally, I would like to give a big thanks to the Publication Chairs, Joakim Nivre and Noah A. Smith, who were very organized and handily managed the job of pulling all materials together for the main conference and workshop proceedings, no small feat.

In addition to the Chairs, individuals within the ACL organization itself deserve recognition. First and foremost, my thanks goes to Dragomir Radev, who provided guidance about what to do next at every step and who had the answer to every question I had within seconds. Owen Rambow also provided much needed advice from the perspective of the North American Chapter. Priscilla Rasmussen is critical to the running of the conference, with her organizational history of how things work. Finally, I would like to thank the Coordinating Committee for being available for discussion and for providing advice.

Kathleen McKeown
ACL-08: HLT General Chair

Preface: Program Chairs

The program for ACL-08: HLT features a wide variety of avenues for authors to present their latest work in computational linguistics, information retrieval, and speech technology. The program includes: full papers, short papers, posters, demonstrations, and a student research workshop, as well as pre- and post-conference tutorials and workshops. In our program design, we attempted to combine the successful approach of ACL07, which had four parallel oral sessions of 25-min full paper presentations, with the HLT model of presenting late-breaking results in parallel sessions of 15-min short paper presentations. We also experimented with an idea adopted from Interspeech, in which authors can choose their desired mode of presentation, oral or poster, based on their assessment of how best to present their work. There is no distinction between posters and oral presentations in terms of quality or in terms of how they appear in the Proceedings. Although it will take more than one year to see this change fully taken up by the membership, we were happy to see some authors choose the poster option from the very outset. Area chairs also used their discretion in indicating which submissions would benefit from which mode of presentation. If the number of submissions continues to grow as it has done in the past few years, poster sessions will be one way to managing this growth without creating a large number of parallel sessions.

This year, the program committee received yet another record-breaking number of submissions, with 470 full and 275 short paper submissions. Full papers were due in mid-January, and the program committee accepted 119 (25%) of these, 95 as oral presentations and 24 as posters. Short papers were due in mid-March, and the committee accepted 64 (23%) of these, 32 for oral presentation and 32 for poster presentation.

First and foremost, we thank all the authors for submitting papers describing their recent work; the sheer number of submissions reflects how active our field is. We are greatly indebted to the 34 area chairs who recruited 720 reviewers, and who managed the reviewing process of both full and short papers in their areas. Reviewers wrote three reviews for each full paper submission, and two reviews for each short paper submission, for a staggering total of just under 2000 reviews! Miraculously, there were only a handful of late reviews. Well done everyone!

As the number of submissions and, consequently the number of area chairs, has risen over the last few years, the ACL program committee has moved away from having a face-to-face meeting of all area chairs. For ACL08: HLT, two of the program co-chairs met for two days at Edinburgh University, using email and teleconferencing to get input from the two program co-chairs not based in Europe, and all of the area chairs. For short paper decision making, three of the four program co-chairs held a teleconference, with input from the fourth co-chair by email as time zone differences permitted.

Another first this year was our decision to award several outstanding paper prizes, rather than trying to identify a single best paper. We did this because we felt that it is typical for conferences as large as this to have several particularly exciting, innovative, and well-crafted papers, and it is extremely difficult to compare quality across areas. We asked area chairs to nominate papers for the various awards and then formed an Outstanding Paper Committee, who wish to remain anonymous, and to whom we owe a great debt of gratitude for their hard work at short notice.

As usual, the main program will run for three days: there will be four parallel sessions of paper presentations. One of these is devoted to the Student Research Workshop, which we would like to thank Ebru Abrisoy, Wolfgang Maier and Keisuke Inoue for organizing. There will also be a poster session on Monday evening, with food and drink to keep everyone going. The demo session, organized by Jimmy Lin, will be held concurrently with the poster session. This year there will be five plenary sessions: two for our very distinguished invited speakers, Susan Dumais and Marc Swerts, one for presentation of the four outstanding papers, one for the presentation by this year's Lifetime Achievement Award winner, and finally one for the ACL business meeting.

Also as usual the conference is flanked by tutorial sessions and workshops. We would like to thank Ani Nenkova, Marilyn Walker and Eugene Agichtein for organizing the tutorials, and Ming Zhou, ChengXiang Zhai and Helen Meng for compiling an excellent program of workshops.

We also thank Kathy McKeown, General Conference Chair, the Local Arrangements Committee headed by Chris Brew, the ACL executive committee, for their help and advice, and last year's co-chairs, Antal van den Bosch and Annie Zaenen, for sharing their experience.

Finally, there were three things that made this all possible. First, we were helped immensely by Jason Eisner, who has compiled an excellent 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). This saved us more than once! Second, we employed a recent PhD, James Clarke, to help us get started with START, and to simply deal with the large volume of work that must be processed within the first few days after submissions are received. James kept us sane. Third, there is the invaluable START system for managing paper submission, reviewing, and decision making. We owe Rich Gerber and the START team a million thanks for responding to questions quickly, and even modifying START overnight to provide what we asked for.

Our most sincere thanks go to Joakim Nivre and Noah A. Smith who took all of our labors and put together the wonderful Proceedings you are now reading.

We hope you enjoy the conference,

Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui
ACL-08: HLT Program Chairs

Organizers

General Chair:

Kathleen McKeown, Columbia University, USA

Local Arrangements Chair:

Chris Brew, The Ohio State University, USA

Program Chairs:

Johanna Moore (Natural Language Processing), University of Edinburgh, UK
Simone Teufel (Natural Language Processing), University of Cambridge, UK
James Allan (Information Retrieval), University of Massachusetts, USA
Sadaoki Furui (Speech), Tokyo Institute of Technology, Japan

Student Research Workshop:

Ebru Arisoy (Speech co-chair), Bogazici University, Turkey
Wolfgang Maier (Natural Language Processing co-chair), University of Tuebingen, Germany
Keisuke Inoue (Information Retrieval co-chair), Syracuse University, USA
Janyce Wiebe (Faculty Advisor), University of Pittsburgh, USA

Workshop Chair:

Ming Zhou, Microsoft Research China, China

Tutorial Chairs:

Ani Nenkova (Coordinator), University of Pennsylvania, USA
Marilyn Walker, University of Sheffield, UK
Eugene Agichtein, Emory University, USA

Demo Chair:

Jimmy Lin, University of Maryland, USA

Sponsorship Chairs:

Inderjeet Mani, Mitre Corporation, USA
Josef van Genabith, Dublin City University, Ireland
Michael White, The Ohio State University, USA

Publications Chairs:

Joakim Nivre, Växjö University and Uppsala University, Sweden
Noah Smith, Carnegie Mellon University, USA

Publicity Chairs:

Hal Daumé III, University of Utah, USA
Eric Fosler-Lussier, The Ohio State University, USA
Diane Kelly, University of North Carolina, USA

Student Volunteers:

Ilana Bromberg (Volunteer co-ordinator)
Crystal Nakatsu (Accomodation requests)
Dominic Espinosa (Conference booklet)

Webmaster:

DJ Hovermale, The Ohio State University, USA

Publications Committee:

Marco Kuhlmann, Uppsala University, Sweden
Carol Sisson, Carnegie Mellon University, USA
Filip Salomonsson, Uppsala University, Sweden

Registration:

Priscilla Rasmussen, Association for Computational Linguistics (ACL)

ACL Coordinating Committee:

Nicoletta Calzolari, Universitá di Pisa Cantara, Italy
Jennifer Chu-Carroll, IBM, USA
Graeme Hirst, University of Toronto, Canada
Chris Manning, Stanford University, USA
Kathleen McCoy, University of Delaware, US
Dragomir Radev, University of Michigan, USA
Owen Rambow, Columbia University, USA
Priscilla Rasmussen, Association for Computational Linguistics (ACL)
Mark Steedman, The University of Edinburgh, UK
Suzanne Stevenson, University of Toronto, Canada

Program Committee

Program Chairs:

Johanna D. Moore, University of Edinburgh (UK)
Simone Teufel, Cambridge University (UK)
James Allan, University of Massachusetts Amherst (USA)
Sadaoki Furui, Tokyo Institute of Technology (Japan)

Area Chairs:

Jason Baldridge, University of Texas at Austin (USA)
Regina Barzilay, Massachusetts Institute of Technology (USA)
Pushpak Bhattacharayya, Indian Institute of Technology Bombay (India)
David Carmel, IBM Research (Israel)
David Chiang, USC/Information Sciences Institute (USA)
Steve Clark, Oxford University (UK)
Hal Daumé III, University of Utah (USA)
Dina Demner-Fushman, National Library of Medicine (USA)
Li Deng, Microsoft Research (USA)
Mark Dras, Macquarie University (Australia)
Pascale Fung, Hong Kong University of Science and Technology (China)
Daniel Gildea, University of Rochester (USA)
John Hansen, University of Texas at Dallas (USA)
Daniel Hardt, Copenhagen Business School (Denmark)
Masato Ishizaki, University of Tokyo (Japan)
Michael Johnston, AT&T Labs Reserach (USA)
Min-Yen Kan, National University of Singapore (Singapore)
Noriko Kando, National Institute of Informatics (Japan)
Emiel Krahmer, Tilburg University (Netherlands)
Elizabeth Liddy, Syracuse University (USA)
Chin-Yew Lin, Microsoft Research Asia (China)
Andrew McCallum, University of Massachusetts Amherst (USA)
Katja Markert, University of Leeds (UK)
Lluís Màrquez, Universitat Politecnica de Catalunya (Spain)
Raymond Mooney, University of Texas at Austin (USA)
Rashmi Prasad, University of Pennsylvania (USA)
Helmut Schmid, University of Stuttgart (Germany)
Sabine Schulte im Walde, University of Stuttgart (Germany)
Rohini Srihari, University of Buffalo (USA)
Manfred Stede, Potsdam University (Germany)
Keiichi Tokuda, Nagoya Institute of Technology (Japan)
Taro Watanabe, NTT Communication Science Laboratories (Japan)
Janyce Wiebe, University of Pittsburgh (USA)
David Weir, Sussex University (UK)

Program Committee Members:

Doug Appelt, Steven Abney, Meni Adler, Stergos Afantenos, Eugene Agichtein, Eneko Agirre, Lars Ahrenberg, Salah Ait-Mokhtar, Ahmet Aker, Jan Alexandersson, Afra Alishahi, Yasemin Altun, Sophia Ananiadou, Galen Andrew, Masahiro Araki, Masayuki Asahara, Nicholas Asher, Michaela Atterer, Necip Fazil Ayan

Timothy Baldwin, Srinivas Bangalore, Michele Banko, Colin Bannard, Roy Bar-Haim, Marco Baroni, Roberto Basili, John Bateman, Johnathan Baxter, Tilman Becker, Ron Bekkerman, Anja Belz, José Benedí, Paul Bennett, Sabine Bergler, Kay Berkling, Yves Bestgen, Rahul Bhagat, Indrajit Bhattacharya, Tanmay Bhattacharya, Pushpak Bhattacharyya, Chris Biemann, Dan Bikel, Mikhail Bilenko, Jeff Bilmes, Philippe Blache, Patrick Blackburn, Sasha Blair-Goldensohn, David Blei, John Blitzer, Phil Blunsom, Gemma Boleda, Johan Bos, Pierre Boullier, Karl Branting, Thorsten Brants, Eric Breck, Chris Brew, Ted Briscoe, Chris Brockett, Ralf Brown, Paul Buitelaar, Razvan Bunescu, Harry Bunt, Stephan Busemann, Donna Byron

Aoife Cahill, Charles Callaway, Chris Callison-Burch, Nicoletta Calzolari, Nick Campbell, Yunbo Cao, Sandra Carberry, Giuseppe Carenini, Jean Carletta, David Carmel, Xavier Carreras, John Carroll, Francisco Casacuberta, Justine Cassell, Lawrence Cavedon, Suleyman Cetintas, Yee Seng Chan, Raman Chandrasekar, Jason Chang, Eugene Charniak, Wanxiang Che, Ciprian Chelba, Hsin-Hsi Chen, John Chen, Colin Cherry, David Chiang, Christian Chiarcos, Yejin Choi, Min Chu, Tat-Seng Chua, Jennifer Chu-Carroll, Ken Church, Massimiliano Ciaramita, Philip Cimiano, Ariel Cohen, Trevor Cohn, Michael Collins, Alistair Conkie, John Conroy, Robin Cooper, Bonaventura Coppola, Mark Core, Marta Costa-jussá, Koby Crammer, Mark Craven, Josep Crego, Silviu Cucerzan, Hang Cui, Aron Culotta, James Curran

Walter Daelemans, Ido Dagan, Robert Dale, Hoa Dang, Hal Daumé III, Eric de la Clergerie, Maarten de Rijke, Vera Demberg, Dina Demner-Fushman, Yasuharu Den, Steve DeNeefe, John DeNero, Li Deng, Yonggang Deng, Ann Devitt, Barbara di Eugenio, Mona Diab, Fernando Diaz, Anne Diekema, Giuseppe DiFabbrizio, Kohji Dohsaka, Bill Dolan, Bonnie Dorr, John Dowding, Mark Dras, Mark Dredze, Gregory Druck, Amit Dubey, Kevin Duh

Phil Edmonds, Markus Egg, Patrick Ehlen, Andreas Eisele, Jacob Eisenstein, Michael Elhadad, Micha Elsner, Katrin Erk, Gunes Erkan, David Evans, Stefan Evert

Yi Fang, Afsaneh Fazly, Ronen Feldman, Christiane Fellbaum, Raul Fernandez, Elena Filatova, Jenny Finkel, Michael Fleischman, Dan Flickinger, Radu Florian, Katherine Forbes, Eric Fosler-Lussier, Frederik Fouvry, Nissim Francez, Robert Frank, Alex Fraser, Bob Frederking, Marjorie Freedman, Dayne Freitag, Junichi Fukumoto

Evgeniy Gabrilovich, Robert Gaizauskas, Michael Gamon, Sudeep Gandhe, Yuqing Gao, Claire Gardent, Ulrich Germann, Roxana Girju, Natalie Glance, Oren Glickman, Amir Globerson, Yoav Goldberg, Ayelet Goldstein, Jade Goldstein, Sharon Goldwater, Gregory Grefenstette, Thomas Griffiths, Ralph Grishman, Iryna Gurevych, Joakim Gustafson

Stephanie Haas, Nizar Habash, Aria Haghighi, Tom Hain, Dilek Hakkani-Tur, Keith Hall, Sanda Harabagiu, Donna Harman, Sasa Hasan, Timothy Hazen, Daqing He, Xiaodong He, Mary Hearne, Marti Hearst, Ulrich Heid, James Henderson, Ulf Hermjakob, Andrew Hickl, Julia Hirschberg, Lynette Hirschman, Graeme Hirst, Julia Hockenmaier, Mark Hopkins, Veronique Hoste, Eduard Hovy, Churen Huang, Liang Huang, Sarmad Hussain, Bouke Huurnink, Mei-Yuh Hwang

Nancy Ide, Diana Inkpen, Kentaro Inui, Mitsuru Ishizuka, Abe Ittycheriah

Jagadeesh Jagarlamudi, Martin Jansche, Mark Johnson, Rie Johnson, Kristiina Jokinen, Gareth Jones, Rosie Jones, Aravind Joshi

Laura Kallmeyer, Nanda Kambhatla, Hiroshi Kanayama, Noriko Kando, Damianos Karakos, Nikiforos Karamanis, Hideki Kashioka, Yasuhiro Katagiri, Rohit Kate, Tsuneaki Kato, Boris Katz, Tatsuya Kawahara, Junichi Kazama, Bill Keler, Frank Keller, Charles Kemp, Andre Kempe, Stanley Yong Wai Keong, Sharam Khadivi, Mahboob Khalid, Rodger Kibble, Bernd Kiefer, Adam Kilgarriff, Chunyu Kit, Dan Klein, Kevin Knight, Alistair Knott, Philipp Koehn, Rob Koeling, Alexander Koller, Terry Koo, Moshe Koppel, Anna Korhonen, Kimmo Koskenniemi, Emiel Krahmer, Geert-Jan Kruijff, Yuval Krymlowski, Sandra Kuebler, Marco Kuhlmann, Jonas Kuhn, Seth Kulick, Shankar Kumar, A Kumaran, June-Jei Kuo, Sadao Kurohashi

Philippe Langlais, Mirella Lapata, Alex Lascarides, Alberto Lavelli, Alon Lavie, Victor Lavrenko, Alan Lee, Gary Lee, Lillian Lee, Yoong Keok Lee, Xin Lei, Gregor Leusch, Lori Levin, Hang Li, Jianguo Li, Qing Li, Xiaolong Li, Xiaoyan Li, Jimmy Lin, Krister Lindén, Lucian Lita, Ken Litkowski, Diane Litman, Bing Liu, Qun Liu, Tie-Yan Liu, Yang Liu, Karen Livescu, Andrei Ljolje, Adam Lopez, Yajuan Lu, Anke Lüdeling, Xiaoqiang Luo

Brian Mak, Rob Malouf, Inderjeet Mani, Gideon Mann, Daniel Marcu, Lluís Màrquez, Brandeis Marshall, Maria Antònia Martí, James Martin, Jean-Claude Martin, David Martínez, Gregory Marton, Mstislav Maslennikov, Tomoko Matsui, Yuji Matsumoto, Evgeny Matusov, Arne Mauser, Jonathan May, Mark Maybury, Diana McCarthy, Mark McConnville, Kathleen McCoy, Ryan McDonald, Tony Mcenry, Chris Mellish, Helen Meng, Paola Merlo, Detmar Meurers, Rada Mihalcea, Brian Milch, Eleni Miltsakaki, David Mimno, Wolfgang Minker, Einat Minkov, Gilad Mishne, Dipti Misra, Teruko Mitamura, Mandar Mitra, Vibhu Mittal, Yusuke Miyao, Noboru Miyazaki, Sien Moens, Saif Mohammad, Rajat Mohanty, Dan Moldovan, Diego Mollá, Christian Monson, Christof Monz, Raymond Mooney, Bob Moore, Glyn Morrill, Alessandro Moschitti, Karin Müller, Dragos Munteanu, Smaranda Muresan, Reinhard Muskens, Sung Hyon Myaeng

Masaaki Nagata, Mikio Nakano, Yukiko Nakano, Vivi Nastase, Roberto Navigli, Mark-Jan Nederhof, Ani Nenkova, John Nerbonne, Günter Neumann, Hermann Ney, Hwee Tou Ng, Vincent Ng, Patrick Nguyen, Jian-Yun Nie, Zaiqing Nie, Takashi Ninomiya, Malvina Nissim, Cheng Niu, Joakim Nivre, Chikashi Nobata, Elena Not, Adrian Novischi

Jon Oberlander, Franz Och, Stephan Oepen, Kemal Oflazer, Manabu Okumura, Miles Osborne, Jahna Otterbacher

Sebastian Pado, Tim Paek, Martha Palmer, Bo Pang, Cecile Paris, Marius Pasca, Rebecca Passonneau, Jon Patrick, Siddharth Patwardhan, Michael Paul, Adam Pease, Ted Pedersen, Catherine Pelachaud, Anselmo Peñas, Gerald Penn, Wim Peters, Paul Piwek, Massimo Poesio, Octavian Popescu, Andrei Popescu-Belis, Maja Popovic, Chris Potts, Richard Power, Sameer Pradhan, John Prager, Rashmi Prasad, Detlef Prescher, Stephen Pulman, Amruta Purandare, James Pustejovsky

Long Qiu, Yan Qu, Chris Quirk

Hema Raghavan, Bhuvana Ramabhadran, Ganesh Ramakrishnan, Owen Rambow, Lance Ramshaw, Deepak Ravichandran, Ehud Reiter, Norbert Reithinger, Philip Resnik, Giuseppe Riccardi, Stefan Riezler, German Rigau, Ellen Riloff, Hae-Chang Rim, Fabio Rinaldi, Brian Roark, James Rogers, Maribel Romero, Barbara Rosario, Dan Roth, Victoria Rubin

Kenji Sagae, Horacio Saggion, Tetsuya Sakai, Joan A. Sánchez, Mark Sanderson, Murat Saraclar, Anoop Sarkar, Shudeshna Sarkar, Yutaka Sasaki, Giorgio Satta, Jan Schehl, Michael Schiehlen, Anne Schiller, David Schlangen, Judith Schlesinger, Helmut Schmid, Marc Schroeder, Hinrich Schütze, Holger Schwenk, Donia Scott, Satoshi Sekine, Mike Seltzer, Vijay Shanker, Libin Shen, Akira Shimazu, Luo Si, Advaith Siddharthan, Melanie Siegel, Khalil Simaan, Michel Simard, David Smith, Rion Snow, Benjamin Snyder, Stephen Soderland, Anders Søgaard, Swapna Somasundaran, David Sontag, Jennifer Spenader, Caroline Sporleder, Richard Sproat, Manfred Stede, Mark Steedman, Amanda Stent, Mark Stevenson, Suzanne Stevenson, Nicola Stokes, Matthew Stone, Veselin Stoyanov, Carlo Strapparava, Michael Strube, Tomek Strzalkowski, Jian Su, Keh-Yih Su, Eiichiro Sumita, Jian-Tao Sun, Richard Sutcliffe, Charles Sutton, Idan Szpektor

Maite Taboada, John Tait, Hiroya Takamura, David Talbot, Pasi Tapanainen, Joel Tetreault, Mariet Theune, Vu Thuy, Jörg Tiedemann, Christoph Tillmann, Roberto Togneri, Takenobu Tokunaga, Kristina Toutanova, David Traum, Benjamin Tsou, Hajime Tsukada, Yoshimasa Tsuruoka, Gokhan Tur, Peter Turney

Raghavendra Udupa, Nicola Ueffing, Masao Utiyama

Antal van den Bosch, Josef van Genabith, Hans van Halteren, Lucy Vanderwende, Tony Veale, Sriram Venkatapathy, Ashish Venugopal, Marc Verhagen, Paola Verlardi, Yannick Versley, Renata Vieira, David Vilar, Piek Vossen, Atro Voutilainen

Joachim Wagner, Marilyn Walker, Michael Walsh, Xiaojun Wan, Haifeng Wang, Wei Wang, Bonnie Webber, Wouter Weerkamp, Ben Wellner, Fuiliang Weng, Michael White, Richard Wicentowski, Yorick Wilks, Theresa Wilson, Shuly Wintner, Yuk Wah Wong, Johan Wouters, Dekai Wu

Fei Xia, Jingfang Xu, Peng Xu

Atsushi Yamada, Kazuhide Yamamoto, Xiaofeng Yang, Alexander Yates, Shiren Ye, Scott Wen-tau Yih, Clem Yu, Deniz Yuret

Dmitry Zaykovskiy, Dmitry Zelenko, Luke Zettlemoyer, ChengXiang Zhai, Hao Zhang, Min Zhang, Rong Zhang, Tong Zhang, Yue Zhang, Jerry Zhu, Andreas Zollmann, Chengqing Zong, Ingrid Zukerman, Pierre Zweigenbaum

Conference Program

Monday, June 16, 2008

9:00–9:10 Opening Session

9:10–10:10 Invited Talk: Marc Swerts, *Facial Expressions in Human-Human and Human-Machine Interactions*

10:10–10:40 **Break**

Session 1A: Information Extraction 1

10:40–11:05 *Mining Wiki Resources for Multilingual Named Entity Recognition*
Alexander E. Richman and Patrick Schone

11:05–11:30 *Distributional Identification of Non-Referential Pronouns*
Shane Bergsma, Dekang Lin and Randy Goebel

11:30–11:55 *Weakly-Supervised Acquisition of Open-Domain Classes and Class Attributes from Web Documents and Query Logs*
Marius Paşca and Benjamin Van Durme

11:55–12:20 *The Tradeoffs Between Open and Traditional Relation Extraction*
Michele Banko and Oren Etzioni

Session 1B: Language Resources and Evaluation

10:40–11:05 *PDT 2.0 Requirements on a Query Language*
Jiří Mírovský

11:05–11:30 *Task-oriented Evaluation of Syntactic Parsers and Their Representations*
Yusuke Miyao, Rune Sætre, Kenji Sagae, Takuya Matsuzaki and Jun'ichi Tsujii

11:30–11:55 *MAXSIM: A Maximum Similarity Metric for Machine Translation Evaluation*
Yee Seng Chan and Hwee Tou Ng

11:55–12:20 *Contradictions and Justifications: Extensions to the Textual Entailment Task*
Ellen M. Voorhees

Monday, June 16, 2008 (continued)

Session 1C: Machine Translation 1

10:40–11:05 *Cohesive Phrase-Based Decoding for Statistical Machine Translation*
Colin Cherry

11:05–11:30 *Phrase Table Training for Precision and Recall: What Makes a Good Phrase and a Good Phrase Pair?*
Yonggang Deng, Jia Xu and Yuqing Gao

11:30–11:55 *Measure Word Generation for English-Chinese SMT Systems*
Dongdong Zhang, Mu Li, Nan Duan, Chi-Ho Li and Ming Zhou

11:55–12:20 *Bayesian Learning of Non-Compositional Phrases with Synchronous Parsing*
Hao Zhang, Chris Quirk, Robert C. Moore and Daniel Gildea

Session 1D: Speech Processing

10:40–11:05 *Applying a Grammar-Based Language Model to a Simplified Broadcast-News Transcription Task*
Tobias Kaufmann and Beat Pfister

11:05–11:30 *Automatic Editing in a Back-End Speech-to-Text System*
Maximilian Bisani, Paul Vozila, Olivier Divay and Jeff Adams

11:30–11:55 *Grounded Language Modeling for Automatic Speech Recognition of Sports Video*
Michael Fleischman and Deb Roy

11:55–12:20 *Lexicalized Phonotactic Word Segmentation*
Margaret M. Fleck

12:20–2:00 **Lunch**

Monday, June 16, 2008 (continued)

Session 2A: Information Retrieval 1

2:00–2:25 *A Re-examination of Query Expansion Using Lexical Resources*
Hui Fang

2:25–2:50 *Selecting Query Term Alternations for Web Search by Exploiting Query Contexts*
Guihong Cao, Stephen Robertson and Jian-Yun Nie

2:50–3:15 *Searching Questions by Identifying Question Topic and Question Focus*
Huizhong Duan, Yunbo Cao, Chin-Yew Lin and Yong Yu

Session 2B: Language Generation

2:00–2:25 *Trainable Generation of Big-Five Personality Styles through Data-Driven Parameter Estimation*
François Mairesse and Marilyn Walker

2:25–2:50 *Correcting Misuse of Verb Forms*
John Lee and Stephanie Seneff

2:50–3:15 *Hypertagging: Supertagging for Surface Realization with CCG*
Dominic Espinosa, Michael White and Dennis Mehay

Session 2C: Machine Translation 2

2:00–2:25 *Forest-Based Translation*
Haitao Mi, Liang Huang and Qun Liu

2:25–2:50 *A Discriminative Latent Variable Model for Statistical Machine Translation*
Phil Blunsom, Trevor Cohn and Miles Osborne

2:50–3:15 *Efficient Multi-Pass Decoding for Synchronous Context Free Grammars*
Hao Zhang and Daniel Gildea

Monday, June 16, 2008 (continued)

Session 2D: Semantics 1

2:00–2:25 *Regular Tree Grammars as a Formalism for Scope Underspecification*
Alexander Koller, Michaela Regneri and Stefan Thater

2:25–2:50 *Classification of Semantic Relationships between Nominals Using Pattern Clusters*
Dmitry Davidov and Ari Rappoport

2:50–3:15 *Vector-based Models of Semantic Composition*
Jeff Mitchell and Mirella Lapata

3:15–3:45 **Break**

Session 3A: Information Extraction 2

3:45–4:10 *Exploiting Feature Hierarchy for Transfer Learning in Named Entity Recognition*
Andrew Arnold, Ramesh Nallapati and William W. Cohen

4:10–4:35 *Refining Event Extraction through Cross-Document Inference*
Heng Ji and Ralph Grishman

4:35–5:00 *Learning Document-Level Semantic Properties from Free-Text Annotations*
S.R.K. Branavan, Harr Chen, Jacob Eisenstein and Regina Barzilay

5:00–5:25 *Automatic Image Annotation Using Auxiliary Text Information*
Yansong Feng and Mirella Lapata

Monday, June 16, 2008 (continued)

Session 3B: Sentiment Analysis

3:45–4:10 *Hedge Classification in Biomedical Texts with a Weakly Supervised Selection of Keywords*
György Szarvas

4:10–4:35 *When Specialists and Generalists Work Together: Overcoming Domain Dependence in Sentiment Tagging*
Alina Andreevskaia and Sabine Bergler

4:35–5:00 *A Generic Sentence Trimmer with CRFs*
Tadashi Nomoto

5:00–5:25 *A Joint Model of Text and Aspect Ratings for Sentiment Summarization*
Ivan Titov and Ryan McDonald

Session 3C: Syntax and Parsing 1

3:45–4:10 *Improving Parsing and PP Attachment Performance with Sense Information*
Eneko Agirre, Timothy Baldwin and David Martinez

4:10–4:35 *A Logical Basis for the D Combinator and Normal Form in CCG*
Frederick Hoyt and Jason Baldridge

4:35–5:00 *Parsing Noun Phrase Structure with CCG*
David Vadas and James R. Curran

5:00–5:25 *Sentence Simplification for Semantic Role Labeling*
David Vickrey and Daphne Koller

Monday, June 16, 2008 (continued)

Session 3D: Student Research Workshop

3:45–4:10 *A Supervised Learning Approach to Automatic Synonym Identification Based on Distributional Features*
Masato Hagiwara

4:10–4:35 *An Integraged Architecture for Generating Parenthetical Constructions*
Eva Banik

4:35–5:00 *Inferring Activity Time in News through Event Modeling*
Vladimir Eidelman

5:00–5:25 *Combining Source and Target Language Information for Name Tagging of Machine Translation Output*
Shasha Liao

5:25–5:50 *A Re-examination on Features in Regression Based Approach to Automatic MT Evaluation*
Shuqi Sun, Yin Chen and Jufeng Li

5:25–6:00 **Break**

6:00–8:30 **Poster and Demo Session**

Long Paper Posters

Summarizing Emails with Conversational Cohesion and Subjectivity
Giuseppe Carenini, Raymond T. Ng and Xiaodong Zhou

Ad Hoc Treebank Structures
Markus Dickinson

A Single Generative Model for Joint Morphological Segmentation and Syntactic Parsing
Yoav Goldberg and Reut Tsarfaty

Which Words Are Hard to Recognize? Prosodic, Lexical, and Disfluency Factors that Increase ASR Error Rates
Sharon Goldwater, Dan Jurafsky and Christopher D. Manning

Name Translation in Statistical Machine Translation - Learning When to Transliterate
Ulf Hermjakob, Kevin Knight and Hal Daumé III

Using Adaptor Grammars to Identify Synergies in the Unsupervised Acquisition of Linguistic Structure
Mark Johnson

Inducing Gazetteers for Named Entity Recognition by Large-Scale Clustering of Dependency Relations
Jun'ichi Kazama and Kentaro Torisawa

Evaluating Roget's Thesauri
Alistair Kennedy and Stan Szpakowicz

Unsupervised Translation Induction for Chinese Abbreviations using Monolingual Corpora
Zhifei Li and David Yarowsky

Which Are the Best Features for Automatic Verb Classification
Jianguo Li and Chris Brew

Collecting a Why-Question Corpus for Development and Evaluation of an Automatic QA-System
Joanna Mrozinski, Edward Whittaker and Sadaoki Furui

Solving Relational Similarity Problems Using the Web as a Corpus
Preslav Nakov and Marti A. Hearst

Combining Speech Retrieval Results with Generalized Additive Models
J. Scott Olsson and Douglas W. Oard

A Critical Reassessment of Evaluation Baselines for Speech Summarization
Gerald Penn and Xiaodan Zhu

Intensional Summaries as Cooperative Responses in Dialogue: Automation and Evaluation
Joseph Polifroni and Marilyn Walker

Word Clustering and Word Selection Based Feature Reduction for MaxEnt Based Hindi NER
Sujan Kumar Saha, Pabitra Mitra and Sudeshna Sarkar

Combining EM Training and the MDL Principle for an Automatic Verb Classification Incorporating Selectional Preferences
Sabine Schulte im Walde, Christian Hying, Christian Scheible and Helmut Schmid

Randomized Language Models via Perfect Hash Functions
David Talbot and Thorsten Brants

Applying Morphology Generation Models to Machine Translation
Kristina Toutanova, Hisami Suzuki and Achim Ruopp

Multilingual Harvesting of Cross-Cultural Stereotypes
Tony Veale, Yanfen Hao and Guofu Li

Semi-Supervised Convex Training for Dependency Parsing
Qin Iris Wang, Dale Schuurmans and Dekang Lin

Chinese-English Backward Transliteration Assisted with Mining Monolingual Web Pages
Fan Yang, Jun Zhao, Bo Zou, Kang Liu and Feifan Liu

Robustness and Generalization of Role Sets: PropBank vs. VerbNet
Beñat Zapirain, Eneko Agirre and Lluís Màrquez

A Tree Sequence Alignment-based Tree-to-Tree Translation Model
Min Zhang, Hongfei Jiang, Aiti Aw, Haizhou Li, Chew Lim Tan and Sheng Li

Short Paper Posters

Language Dynamics and Capitalization using Maximum Entropy
Fernando Batista, Nuno Mamede and Isabel Trancoso

Surprising Parser Actions and Reading Difficulty
Marisa Ferrara Boston, John T. Hale, Reinhold Kliegl and Shravan Vasishth

Improving the Performance of the Random Walk Model for Answering Complex Questions
Yllias Chali and Shafiq Joty

Dimensions of Subjectivity in Natural Language
Wei Chen

Extractive Summaries for Educational Science Content
Sebastian de la Chica, Faisal Ahmad, James H. Martin and Tamara Sumner

Dialect Classification for Online Podcasts Fusing Acoustic and Language Based Structural and Semantic Information
Rahul Chitturi and John Hansen

The Complexity of Phrase Alignment Problems
John DeNero and Dan Klein

Novel Semantic Features for Verb Sense Disambiguation
Dmitriy Dligach and Martha Palmer

Icelandic Data Driven Part of Speech Tagging
Mark Dredze and Joel Wallenberg

Beyond Log-Linear Models: Boosted Minimum Error Rate Training for N-best Re-ranking
Kevin Duh and Katrin Kirchhoff

Coreference-inspired Coherence Modeling
Micha Elsner and Eugene Charniak

Enforcing Transitivity in Coreference Resolution
Jenny Rose Finkel and Christopher D. Manning

Simulating the Behaviour of Older versus Younger Users when Interacting with Spoken Dialogue Systems
Kallirroi Georgila, Maria Wolters and Johanna Moore

Active Sample Selection for Named Entity Transliteration
Dan Goldwasser and Dan Roth

Four Techniques for Online Handling of Out-of-Vocabulary Words in Arabic-English Statistical Machine Translation
Nizar Habash

Combined One Sense Disambiguation of Abbreviations
Yaakov HaCohen-Kerner, Ariel Kass and Ariel Peretz

Assessing the Costs of Sampling Methods in Active Learning for Annotation
Robbie Haertel, Eric Ringger, Kevin Seppi, Carroll James and McClanahan Peter

Blog Categorization Exploiting Domain Dictionary and Dynamically Estimated Domains of Unknown Words
Chikara Hashimoto and Sadao Kurohashi

Mixture Model POMDPs for Efficient Handling of Uncertainty in Dialogue Management
James Henderson and Oliver Lemon

Recent Improvements in the CMU Large Scale Chinese-English SMT System
Almut Silja Hildebrand, Kay Rottmann, Mohamed Noamany, Quin Gao, Sanjika Hewavitharana, Nguyen Bach and Stephan Vogel

Machine Translation System Combination using ITG-based Alignments
Damianos Karakos, Jason Eisner, Sanjeev Khudanpur and Markus Dreyer

Dictionary Definitions based Homograph Identification using a Generative Hierarchical Model
Anagha Kulkarni and Jamie Callan

A Novel Feature-based Approach to Chinese Entity Relation Extraction
Wenjie Li, Peng Zhang, Furu Wei, Yuexian Hou and Qin Lu

Using Structural Information for Identifying Similar Chinese Characters
Chao-Lin Liu and Jen-Hsiang Lin

You've Got Answers: Towards Personalized Models for Predicting Success in Community Question Answering
Yandong Liu and Eugene Agichtein

Self-Training for Biomedical Parsing
David McClosky and Eugene Charniak

A Unified Syntactic Model for Parsing Fluent and Disfluent Speech
Tim Miller and William Schuler

The Good, the Bad, and the Unknown: Morphosyllabic Sentiment Tagging of Unseen Words
Karo Moilanen and Stephen Pulman

Kernels on Linguistic Structures for Answer Extraction
Alessandro Moschitti and Silvia Quarteroni

Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking
Ryan Roth, Owen Rambow, Nizar Habash, Mona Diab and Cynthia Rudin

Using Automatically Transcribed Dialogs to Learn User Models in a Spoken Dialog System
Umar Syed and Jason Williams

Robust Extraction of Named Entity Including Unfamiliar Word
Masatoshi Tsuchiya, Shinya Hida and Seiichi Nakagawa

In-Browser Summarisation: Generating Elaborative Summaries Biased Towards the Reading Context
Stephen Wan and Cécile Paris

Lyric-based Song Sentiment Classification with Sentiment Vector Space Model
Yunqing Xia, Linlin Wang, Kam-Fai Wong and Mingxing Xu

Mining Wikipedia Revision Histories for Improving Sentence Compression
Elif Yamangil and Rani Nelken

Smoothing a Tera-word Language Model
Deniz Yuret

Student Research Workshop Posters

The Role of Positive Feedback in Intelligent Tutoring Systems
Davide Fossati

Arabic Language Modeling with Finite State Transducers
Ilana Heintz

Impact of Initiative on Collaborative Problem Solving
Cynthia Kersey

An Unsupervised Vector Approach to Biomedical Term Disambiguation: Integrating UMLS and Medline
Bridget McInnes

A Subcategorization Acquisition System for French Verbs
Cédric Messiant

Adaptive Language Modeling for Word Prediction
Keith Trnka

A Hierarchical Approach to Encoding Medical Concepts for Clinical Notes
Yitao Zhang

Demonstrations

Demonstration of a POMDP Voice Dialer
Jason Williams

Generating Research Websites Using Summarisation Techniques
Advaith Siddharthan and Ann Copestake

BART: A Modular Toolkit for Coreference Resolution
Yannick Versley, Simone Paolo Ponzetto, Massimo Poesio, Vladimir Eidelman, Alan Jern, Jason Smith, Xiaofeng Yang and Alessandro Moschitti

Demonstration of the UAM CorpusTool for Text and Image Annotation
Mick O'Donnell

Interactive ASR Error Correction for Touchscreen Devices
David Huggins-Daines and Alexander I. Rudnicky

Yawat: Yet Another Word Alignment Tool
Ulrich Germann

SIDE: The Summarization Integrated Development Environment
Moonyoung Kang, Sourish Chaudhuri, Mahesh Joshi and Carolyn P. Rosé

ModelTalker Voice Recorder—An Interface System for Recording a Corpus of Speech for Synthesis
Debra Yarrington, John Gray, Chris Pennington, H. Timothy Bunnell, Allegra Cornaglia, Jason Lilley, Kyoko Nagao and James Polikoff

The QuALiM Question Answering Demo: Supplementing Answers with Paragraphs drawn from Wikipedia
Michael Kaisser

Tuesday, June 17, 2008

Session: Outstanding Paper Award Presentations

9:00–9:10 Presentation of Awards

9:10–9:35 *Automatic Syllabification with Structured SVMs for Letter-to-Phoneme Conversion*
Susan Bartlett, Grzegorz Kondrak and Colin Cherry

9:35–10:00 *A New String-to-Dependency Machine Translation Algorithm with a Target Dependency Language Model*
Libin Shen, Jinxi Xu and Ralph Weischedel

10:00–10:25 *Forest Reranking: Discriminative Parsing with Non-Local Features*
Liang Huang

10:25–10:40 *Event Matching Using the Transitive Closure of Dependency Relations*
Daniel M. Bikel and Vittorio Castelli

10:40–11:10 **Break**

Session 4A: Syntax and Parsing 2

11:10–11:35 *Simple Semi-supervised Dependency Parsing*
Terry Koo, Xavier Carreras and Michael Collins

11:35–12:00 *Optimal k-arization of Synchronous Tree-Adjoining Grammar*
Rebecca Nesson, Giorgio Satta and Stuart M. Shieber

12:00–12:25 *Enhancing Performance of Lexicalised Grammars*
Rebecca Dridan, Valia Kordoni and Jeremy Nicholson

Tuesday, June 17, 2008 (continued)

Session 4B: Dialogue

11:10–11:35 *Assessing Dialog System User Simulation Evaluation Measures Using Human Judges*
Hua Ai and Diane J. Litman

11:35–12:00 *Robust Dialog Management with N-Best Hypotheses Using Dialog Examples and Agenda*
Cheongjae Lee, Sangkeun Jung and Gary Geunbae Lee

12:00–12:25 *Learning Effective Multimodal Dialogue Strategies from Wizard-of-Oz Data: Bootstrapping and Evaluation*
Verena Rieser and Oliver Lemon

Session 4C: Machine Learning 2

11:10–11:35 *Phrase Chunking Using Entropy Guided Transformation Learning*
Ruy Luiz Milidiú, Cícero Nogueira dos Santos and Julio C. Duarte

11:35–12:00 *Learning Bigrams from Unigrams*
Xiaojin Zhu, Andrew B. Goldberg, Michael Rabbat and Robert Nowak

12:00–12:25 *Semi-Supervised Sequential Labeling and Segmentation Using Giga-Word Scale Unlabeled Data*
Jun Suzuki and Hideki Isozaki

Session 4D: Semantics 2

11:10–11:35 *Large Scale Acquisition of Paraphrases for Learning Surface Patterns*
Rahul Bhagat and Deepak Ravichandran

11:35–12:00 *Contextual Preferences*
Idan Szpektor, Ido Dagan, Roy Bar-Haim and Jacob Goldberger

12:00–12:25 *Unsupervised Discovery of Generic Relationships Using Pattern Clusters and its Evaluation by Automatically Generated SAT Analogy Questions*
Dmitry Davidov and Ari Rappoport

12:25–2:00 **Lunch**

Tuesday, June 17, 2008 (continued)

Session 5A: Short Papers 1 (Machine Translation)

2:00–2:15 *A Linguistically Annotated Reordering Model for BTG-based Statistical Machine Translation*
Deyi Xiong, Min Zhang, Aiti Aw and Haizhou Li

2:15–2:30 *Segmentation for English-to-Arabic Statistical Machine Translation*
Ibrahim Badr, Rabih Zbib and James Glass

2:30–2:45 *Exploiting N-best Hypotheses for SMT Self-Enhancement*
Boxing Chen, Min Zhang, Aiti Aw and Haizhou Li

2:45–3:00 *Partial Matching Strategy for Phrase-based Statistical Machine Translation*
Zhongjun He, Qun Liu and Shouxun Lin

Session 5B: Short Papers 2 (Speech)

2:00–2:15 No presentation

2:15–2:30 *Unsupervised Learning of Acoustic Sub-word Units*
Balakrishnan Varadarajan, Sanjeev Khudanpur and Emmanuel Dupoux

2:30–2:45 *High Frequency Word Entrainment in Spoken Dialogue*
Ani Nenkova, Agustín Gravano and Julia Hirschberg

2:45–3:00 *Distributed Listening: A Parallel Processing Approach to Automatic Speech Recognition*
Yolanda McMillian and Juan Gilbert

Tuesday, June 17, 2008 (continued)

Session 5C: Short Papers 3 (Semantics)

2:00–2:15 *Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations*
Steven Bethard and James H. Martin

2:15–2:30 *Evolving New Lexical Association Measures Using Genetic Programming*
Jan Šnajder, Bojana Dalbelo Bašič, Saša Petrovič and Ivan Sikirič

2:30–2:45 *Semantic Types of Some Generic Relation Arguments: Detection and Evaluation*
Sophia Katrenko and Pieter Adriaans

2:45–3:00 *Mapping between Compositional Semantic Representations and Lexical Semantic Resources: Towards Accurate Deep Semantic Parsing*
Sergio Roa, Valia Kordoni and Yi Zhang

Session 5D: Short Papers 4 (Generation/Summarization)

2:00–2:15 *Query-based Sentence Fusion is Better Defined and Leads to More Preferred Results than Generic Sentence Fusion*
Emiel Krahmer, Erwin Marsi and Paul van Pelt

2:15–2:30 *Intrinsic vs. Extrinsic Evaluation Measures for Referring Expression Generation*
Anja Belz and Albert Gatt

2:30–2:45 *Correlation between ROUGE and Human Evaluation of Extractive Meeting Summaries*
Feifan Liu and Yang Liu

2:45–3:00 *FastSum: Fast and Accurate Query-based Multi-document Summarization*
Frank Schilder and Ravikumar Kondadadi

3:00–3:15 **Break**

Tuesday, June 17, 2008 (continued)

Session 5E: Short Papers 1 (Syntax)

3:15–3:30 *Construct State Modification in the Arabic Treebank*
Ryan Gabbard and Seth Kulick

3:30–3:45 *Unlexicalised Hidden Variable Models of Split Dependency Grammars*
Gabriele Antonio Musillo and Paola Merlo

3:45–4:00 *Computing Confidence Scores for All Sub Parse Trees*
Feng Lin and Fuliang Weng

4:00–4:15 *Adapting a WSJ-Trained Parser to Grammatically Noisy Text*
Jennifer Foster, Joachim Wagner and Josef van Genabith

Session 5F: Short Papers 2 (Dialog/Statistical Methods)

3:15–3:30 *Enriching Spoken Language Translation with Dialog Acts*
Vivek Kumar Rangarajan Sridhar, Srinivas Bangalore and Shrikanth Narayanan

3:30–3:45 *Speakers' Intention Prediction Using Statistics of Multi-level Features in a Schedule Management Domain*
Donghyun Kim, Hyunjung Lee, Choong-Nyoung Seon, Harksoo Kim and Jungyun Seo

3:45–4:00 *Active Learning with Confidence*
Mark Dredze and Koby Crammer

4:00–4:15 *splitSVM: Fast, Space-Efficient, non-Heuristic, Polynomial Kernel Computation for NLP Applications*
Yoav Goldberg and Michael Elhadad

Tuesday, June 17, 2008 (continued)

Session 5G: Short Papers 3 (Semantics/Phonology)

3:15–3:30 *Extracting a Representation from Text for Semantic Analysis*
Rodney D. Nielsen, Wayne Ward, James H. Martin and Martha Palmer

3:30–3:45 *Efficient Processing of Underspecified Discourse Representations*
Michaela Regneri, Markus Egg and Alexander Koller

3:45–4:00 *Choosing Sense Distinctions for WSD: Psycholinguistic Evidence*
Susan Windisch Brown

4:00–4:15 *Decompounding query keywords from compounding languages*
Enrique Alfonseca, Slaven Bilac and Stefan Pharies

Session 5H: Short Papers 4 (Information Retrieval/Sentiment Analysis)

3:15–3:30 *Multi-domain Sentiment Classification*
Shoushan Li and Chengqing Zong

3:30–3:45 *Evaluating Word Prediction: Framing Keystroke Savings*
Keith Trnka and Kathleen McCoy

3:45–4:00 *Pairwise Document Similarity in Large Collections with MapReduce*
Tamer Elsayed, Jimmy Lin and Douglas Oard

4:00–4:15 *Text Segmentation with LDA-Based Fisher Kernel*
Qi Sun, Runxin Li, Dingsheng Luo and Xihong Wu

4:15–4:45 **Break**

Tuesday, June 17, 2008 (continued)

Session 6A: Question Answering

4:45–5:10 *Improving Search Results Quality by Customizing Summary Lengths*
Michael Kaisser, Marti A. Hearst and John B. Lowe

5:10–5:35 *Using Conditional Random Fields to Extract Contexts and Answers of Questions from Online Forums*
Shilin Ding, Gao Cong, Chin-Yew Lin and Xiaoyan Zhu

5:35–6:00 *Learning to Rank Answers on Large Online QA Collections*
Mihai Surdeanu, Massimiliano Ciaramita and Hugo Zaragoza

Session 6B: Phonology, Morphology 1

4:45–5:10 *Unsupervised Lexicon-Based Resolution of Unknown Words for Full Morphological Analysis*
Meni Adler, Yoav Goldberg, David Gabay and Michael Elhadad

5:10–5:35 *Unsupervised Multilingual Learning for Morphological Segmentation*
Benjamin Snyder and Regina Barzilay

5:35–6:00 *EM Can Find Pretty Good HMM POS-Taggers (When Given a Good Start)*
Yoav Goldberg, Meni Adler and Michael Elhadad

Session 6C: Machine Translation 3

4:45–5:10 *Distributed Word Clustering for Large Scale Class-Based Language Modeling in Machine Translation*
Jakob Uszkoreit and Thorsten Brants

5:10–5:35 *Enriching Morphologically Poor Languages for Statistical Machine Translation*
Eleftherios Avramidis and Philipp Koehn

5:35–6:00 *Learning Bilingual Lexicons from Monolingual Corpora*
Aria Haghighi, Percy Liang, Taylor Berg-Kirkpatrick and Dan Klein

Tuesday, June 17, 2008 (continued)

Session 6D: Semantics 3

4:45–5:10 *Pivot Approach for Extracting Paraphrase Patterns from Bilingual Corpora*
Shiqi Zhao, Haifeng Wang, Ting Liu and Sheng Li

5:10–5:35 *Unsupervised Learning of Narrative Event Chains*
Nathanael Chambers and Dan Jurafsky

5:35–6:00 *Semantic Role Labeling Systems for Arabic using Kernel Methods*
Mona Diab, Alessandro Moschitti and Daniele Pighin

7:00–11:00 **Banquet**

Wednesday, June 18, 2008

9:00–10:00 Invited Talk: Susan Dumais, *Supporting Searchers in Searching*

10:00–10:30 **Break**

Session 7A: Summarization

10:30–10:55 *An Unsupervised Approach to Biography Production Using Wikipedia*
Fadi Biadsy, Julia Hirschberg and Elena Filatova

10:55–11:20 *Generating Impact-Based Summaries for Scientific Literature*
Qiaozhu Mei and ChengXiang Zhai

11:20–11:45 *Can You Summarize This? Identifying Correlates of Input Difficulty for Multi-Document Summarization*
Ani Nenkova and Annie Louis

Wednesday, June 18, 2008 (continued)

Session 7B: Discourse and Pragmatics

10:30–10:55 *You Talking to Me? A Corpus and Algorithm for Conversation Disentanglement*
Micha Elsner and Eugene Charniak

10:55–11:20 *An Entity-Mention Model for Coreference Resolution with Inductive Logic Programming*
Xiaofeng Yang, Jian Su, Jun Lang, Chew Lim Tan, Ting Liu and Sheng Li

11:20–11:45 *Gestural Cohesion for Topic Segmentation*
Jacob Eisenstein, Regina Barzilay and Randall Davis

Session 7C: Machine Learning 2

10:30–10:55 *Multi-Task Active Learning for Linguistic Annotations*
Roi Reichart, Katrin Tomanek, Udo Hahn and Ari Rappoport

10:55–11:20 *Generalized Expectation Criteria for Semi-Supervised Learning of Conditional Random Fields*
Gideon S. Mann and Andrew McCallum

11:20–11:45 *Analyzing the Errors of Unsupervised Learning*
Percy Liang and Dan Klein

Session 7D: Phonology, Morphology 2

10:30–10:55 *Joint Word Segmentation and POS Tagging Using a Single Perceptron*
Yue Zhang and Stephen Clark

10:55–11:20 *A Cascaded Linear Model for Joint Chinese Word Segmentation and Part-of-Speech Tagging*
Wenbin Jiang, Liang Huang, Qun Liu and Yajuan Lü

11:20–11:45 *Joint Processing and Discriminative Training for Letter-to-Phoneme Conversion*
Sittichai Jiampojamarn, Colin Cherry and Grzegorz Kondrak

11:45–1:15 **ACL Business Meeting**

1:15–2:30 **Lunch**

Wednesday, June 18, 2008 (continued)

Session 8A: Information Retrieval 2

2:30–2:55 *A Probabilistic Model for Fine-Grained Expert Search*
Shenghua Bao, Huizhong Duan, Qi Zhou, Miao Xiong, Yunbo Cao and Yong Yu

2:55–3:20 *Credibility Improves Topical Blog Post Retrieval*
Wouter Weerkamp and Maarten de Rijke

3:20–3:45 *Linguistically Motivated Features for Enhanced Back-of-the-Book Indexing*
Andras Csomai and Rada Mihalcea

3:45–4:10 *Resolving Personal Names in Email Using Context Expansion*
Tamer Elsayed, Douglas W. Oard and Galileo Namata

Session 8B: Syntax and Parsing 3

2:30–2:55 *Integrating Graph-Based and Transition-Based Dependency Parsers*
Joakim Nivre and Ryan McDonald

2:55–3:20 *Efficient, Feature-based, Conditional Random Field Parsing*
Jenny Rose Finkel, Alex Kleeman and Christopher D. Manning

3:20–3:45 *A Deductive Approach to Dependency Parsing*
Carlos Gómez-Rodríguez, John Carroll and David Weir

3:45–4:10 *Evaluating a Crosslinguistic Grammar Resource: A Case Study of Wambaya*
Emily M. Bender

Wednesday, June 18, 2008 (continued)

Session 8C: Machine Translation 2

2:30–2:55 *Better Alignments = Better Translations?*
Kuzman Ganchev, João V. Graça and Ben Taskar

2:55–3:20 *Mining Parenthetical Translations from the Web by Word Alignment*
Dekang Lin, Shaojun Zhao, Benjamin Van Durme and Marius Paşca

3:20–3:45 *Soft Syntactic Constraints for Hierarchical Phrased-Based Translation*
Yuval Marton and Philip Resnik

3:45–4:10 *Generalizing Word Lattice Translation*
Christopher Dyer, Smaranda Muresan and Philip Resnik

Session 8D: Semantics 4

2:30–2:55 *Combining Multiple Resources to Improve SMT-based Paraphrasing Model*
Shiqi Zhao, Cheng Niu, Ming Zhou, Ting Liu and Sheng Li

2:55–3:20 *Extraction of Entailed Semantic Relations Through Syntax-Based Comma Resolution*
Vivek Srikumar, Roi Reichart, Mark Sammons, Ari Rappoport and Dan Roth

3:20–3:45 *Finding Contradictions in Text*
Marie-Catherine de Marneffe, Anna N. Rafferty and Christopher D. Manning

3:45–4:10 *Semantic Class Learning from the Web with Hyponym Pattern Linkage Graphs*
Zornitsa Kozareva, Ellen Riloff and Eduard Hovy

4:10–4:40 **Break**

4:40–5:50 Lifetime Achievement Award Presentation

5:50–6:00 Closing Session