ACL-08: HLT

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

Short Papers

June 16–17, 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

Table of Contents

Language Dynamics and Capitalization using Maximum Entropy
Fernando Batista, Nuno Mamede and Isabel Trancoso 1

Surprising Parser Actions and Reading Difficulty
Marisa Ferrara Boston, John T. Hale, Reinhold Kliegl and Shravan Vasishth 5

Improving the Performance of the Random Walk Model for Answering Complex Questions
Yllias Chali and Shafiq Joty 9

Dimensions of Subjectivity in Natural Language
Wei Chen 13

Extractive Summaries for Educational Science Content
Sebastian de la Chica, Faisal Ahmad, James H. Martin and Tamara Sumner 17

Dialect Classification for Online Podcasts Fusing Acoustic and Language Based Structural and Semantic Information
Rahul Chitturi and John Hansen 21

The Complexity of Phrase Alignment Problems
John DeNero and Dan Klein 25

Novel Semantic Features for Verb Sense Disambiguation
Dmitriy Dligach and Martha Palmer 29

Icelandic Data Driven Part of Speech Tagging
Mark Dredze and Joel Wallenberg 33

Beyond Log-Linear Models: Boosted Minimum Error Rate Training for N-best Re-ranking
Kevin Duh and Katrin Kirchhoff 37

Coreference-inspired Coherence Modeling
Micha Elsner and Eugene Charniak 41

Enforcing Transitivity in Coreference Resolution
Jenny Rose Finkel and Christopher D. Manning 45

Simulating the Behaviour of Older versus Younger Users when Interacting with Spoken Dialogue Systems
Kallirroi Georgila, Maria Wolters and Johanna Moore 49

Active Sample Selection for Named Entity Transliteration
Dan Goldwasser and Dan Roth 53

Four Techniques for Online Handling of Out-of-Vocabulary Words in Arabic-English Statistical Machine Translation
Nizar Habash 57

Combined One Sense Disambiguation of Abbreviations
Yaakov HaCohen-Kerner, Ariel Kass and Ariel Peretz 61

Assessing the Costs of Sampling Methods in Active Learning for Annotation
Robbie Haertel, Eric Ringger, Kevin Seppi, James Carroll and McClanahan Peter 65

Blog Categorization Exploiting Domain Dictionary and Dynamically Estimated Domains of Unknown Words
Chikara Hashimoto and Sadao Kurohashi 69

Mixture Model POMDPs for Efficient Handling of Uncertainty in Dialogue Management
James Henderson and Oliver Lemon 73

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 77

Machine Translation System Combination using ITG-based Alignments
Damianos Karakos, Jason Eisner, Sanjeev Khudanpur and Markus Dreyer 81

Dictionary Definitions based Homograph Identification using a Generative Hierarchical Model
Anagha Kulkarni and Jamie Callan 85

A Novel Feature-based Approach to Chinese Entity Relation Extraction
Wenjie Li, Peng Zhang, Furu Wei, Yuexian Hou and Qin Lu 89

Using Structural Information for Identifying Similar Chinese Characters
Chao-Lin Liu and Jen-Hsiang Lin 93

You've Got Answers: Towards Personalized Models for Predicting Success in Community Question Answering
Yandong Liu and Eugene Agichtein 97

Self-Training for Biomedical Parsing
David McClosky and Eugene Charniak 101

A Unified Syntactic Model for Parsing Fluent and Disfluent Speech
Tim Miller and William Schuler 105

The Good, the Bad, and the Unknown: Morphosyllabic Sentiment Tagging of Unseen Words
Karo Moilanen and Stephen Pulman 109

Kernels on Linguistic Structures for Answer Extraction
Alessandro Moschitti and Silvia Quarteroni 113

Arabic Morphological Tagging, Diacritization, and Lemmatization Using Lexeme Models and Feature Ranking
Ryan Roth, Owen Rambow, Nizar Habash, Mona Diab and Cynthia Rudin 117

Using Automatically Transcribed Dialogs to Learn User Models in a Spoken Dialog System
Umar Syed and Jason Williams 121

Robust Extraction of Named Entity Including Unfamiliar Word
Masatoshi Tsuchiya, Shinya Hida and Seiichi Nakagawa 125

In-Browser Summarisation: Generating Elaborative Summaries Biased Towards the Reading Context
Stephen Wan and Cécile Paris 129

Lyric-based Song Sentiment Classification with Sentiment Vector Space Model
Yunqing Xia, Linlin Wang, Kam-Fai Wong and Mingxing Xu 133

Mining Wikipedia Revision Histories for Improving Sentence Compression
Elif Yamangil and Rani Nelken 137

Smoothing a Tera-word Language Model
Deniz Yuret 141

Event Matching Using the Transitive Closure of Dependency Relations
Daniel M. Bikel and Vittorio Castelli 145

A Linguistically Annotated Reordering Model for BTG-based Statistical Machine Translation
Deyi Xiong, Min Zhang, Aiti Aw and Haizhou Li 149

Segmentation for English-to-Arabic Statistical Machine Translation
Ibrahim Badr, Rabih Zbib and James Glass 153

Exploiting N-best Hypotheses for SMT Self-Enhancement
Boxing Chen, Min Zhang, Aiti Aw and Haizhou Li 157

Partial Matching Strategy for Phrase-based Statistical Machine Translation
Zhongjun He, Qun Liu and Shouxun Lin 161

Unsupervised Learning of Acoustic Sub-word Units
Balakrishnan Varadarajan, Sanjeev Khudanpur and Emmanuel Dupoux 165

High Frequency Word Entrainment in Spoken Dialogue
Ani Nenkova, Agustín Gravano and Julia Hirschberg 169

Distributed Listening: A Parallel Processing Approach to Automatic Speech Recognition
Yolanda McMillian and Juan Gilbert 173

Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations
Steven Bethard and James H. Martin 177

Evolving New Lexical Association Measures Using Genetic Programming
Jan Šnajder, Bojana Dalbelo Bašić, Saša Petrović and Ivan Sikirić 181

Semantic Types of Some Generic Relation Arguments: Detection and Evaluation
Sophia Katrenko and Pieter Adriaans 185

Mapping between Compositional Semantic Representations and Lexical Semantic Resources: Towards Accurate Deep Semantic Parsing
Sergio Roa, Valia Kordoni and Yi Zhang 189

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 193

Intrinsic vs. Extrinsic Evaluation Measures for Referring Expression Generation
Anja Belz and Albert Gatt 197

Correlation between ROUGE and Human Evaluation of Extractive Meeting Summaries
Feifan Liu and Yang Liu 201

FastSum: Fast and Accurate Query-based Multi-document Summarization
Frank Schilder and Ravikumar Kondadadi 205

Construct State Modification in the Arabic Treebank
Ryan Gabbard and Seth Kulick 209

Unlexicalised Hidden Variable Models of Split Dependency Grammars
Gabriele Antonio Musillo and Paola Merlo 213

Computing Confidence Scores for All Sub Parse Trees
Feng Lin and Fuliang Weng 217

Adapting a WSJ-Trained Parser to Grammatically Noisy Text
Jennifer Foster, Joachim Wagner and Josef van Genabith 221

Enriching Spoken Language Translation with Dialog Acts
Vivek Kumar Rangarajan Sridhar, Srinivas Bangalore and Shrikanth Narayanan 225

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 229

Active Learning with Confidence
Mark Dredze and Koby Crammer 233

splitSVM: Fast, Space-Efficient, non-Heuristic, Polynomial Kernel Computation for NLP Applications
Yoav Goldberg and Michael Elhadad 237

Extracting a Representation from Text for Semantic Analysis
Rodney D. Nielsen, Wayne Ward, James H. Martin and Martha Palmer 241

Efficient Processing of Underspecified Discourse Representations
Michaela Regneri, Markus Egg and Alexander Koller 245

Choosing Sense Distinctions for WSD: Psycholinguistic Evidence
Susan Windisch Brown 249

Decompounding query keywords from compounding languages
Enrique Alfonseca, Slaven Bilac and Stefan Pharies 253

Multi-domain Sentiment Classification
Shoushan Li and Chengqing Zong 257

Evaluating Word Prediction: Framing Keystroke Savings
Keith Trnka and Kathleen McCoy 261

Pairwise Document Similarity in Large Collections with MapReduce
Tamer Elsayed, Jimmy Lin and Douglas Oard 265

Text Segmentation with LDA-Based Fisher Kernel
Qi Sun, Runxin Li, Dingsheng Luo and Xihong Wu 269

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

Monday, June 16, 2008 (continued)

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

Monday, June 16, 2008 (continued)

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, James Carroll 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

10:15–10:30 *Event Matching Using the Transitive Closure of Dependency Relations*
Daniel M. Bikel and Vittorio Castelli

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 Bašić, Saša Petrović and Ivan Sikirić

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