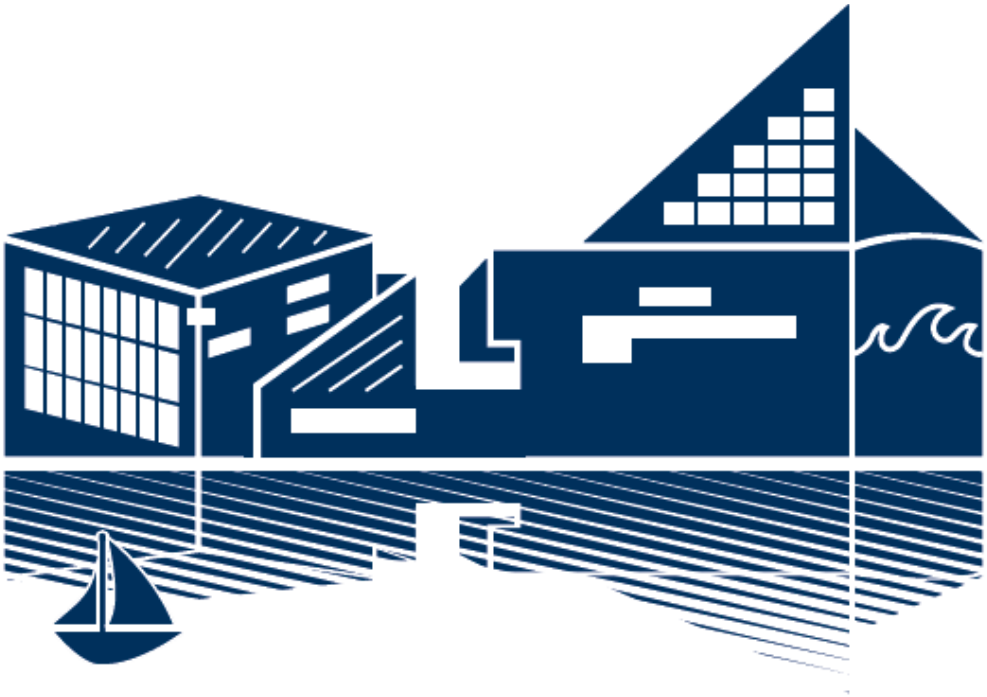

The 52nd Annual Meeting of the
Association for Computational Linguistics

Proceedings of the Conference
Volume 2: Short Papers

ACL 2014
June 22–27
Baltimore

Platinum Level Sponsor:

Gold Level Sponsors:

Silver Level Sponsors:

Bronze Level Sponsors:

Supporters:

©2014 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

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

ISBN 978-1-937284-73-2

Table of Contents

Exploring the Relative Role of Bottom-up and Top-down Information in Phoneme Learning
Abdellah Fourtassi, Thomas Schatz, Balakrishnan Varadarajan and Emmanuel Dupoux 1

Biases in Predicting the Human Language Model
Alex B. Fine, Austin F. Frank, T. Florian Jaeger and Benjamin Van Durme 7

Probabilistic Labeling for Efficient Referential Grounding based on Collaborative Discourse
Changsong Liu, Lanbo She, Rui Fang and Joyce Y. Chai 13

A Composite Kernel Approach for Dialog Topic Tracking with Structured Domain Knowledge from Wikipedia
Seokhwan Kim, Rafael E. Banchs and Haizhou Li 19

An Extension of BLANC to System Mentions
Xiaoqiang Luo, Sameer Pradhan, Marta Recasens and Eduard Hovy 24

Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation
Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng and Michael Strube 30

Measuring Sentiment Annotation Complexity of Text
Aditya Joshi, Abhijit Mishra, Nivvedan Senthamilselvan and Pushpak Bhattacharyya 36

Improving Citation Polarity Classification with Product Reviews
Charles Jochim and Hinrich Schütze 42

Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification
Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou and Ke Xu 49

Sprinkling Topics for Weakly Supervised Text Classification
Swapnil Hingmire and Sutanu Chakraborti 55

A Feature-Enriched Tree Kernel for Relation Extraction
Le Sun and Xianpei Han 61

Employing Word Representations and Regularization for Domain Adaptation of Relation Extraction
Thien Huu Nguyen and Ralph Grishman 68

Graph Ranking for Collective Named Entity Disambiguation
Ayman Alhelbawy and Robert Gaizauskas 75

Descending-Path Convolution Kernel for Syntactic Structures
Chen Lin, Timothy Miller, Alvin Kho, Steven Bethard, Dmitriy Dligach, Sameer Pradhan and Guergana Savova 81

Entities' Sentiment Relevance
Zvi Ben-Ami, Ronen Feldman and Binyamin Rosenfeld 87

Automatic Detection of Multilingual Dictionaries on the Web
Gintare Grigonyte and Timothy Baldwin 93

Automatic Detection of Cognates Using Orthographic Alignment
Alina Maria Ciobanu and Liviu P. Dinu 99

Automatically constructing Wordnet Synsets
Khang Nhut Lam, Feras Al Tarouti and Jugal Kalita 106

Constructing a Turkish-English Parallel TreeBank
Olcay Taner Yıldız, Ercan Solak, Onur Görgün and Razieh Ehsani 112

Improved Typesetting Models for Historical OCR
Taylor Berg-Kirkpatrick and Dan Klein 118

Robust Logistic Regression using Shift Parameters
Julie Tibshirani and Christopher D. Manning 124

Faster Phrase-Based Decoding by Refining Feature State
Kenneth Heafield, Michael Kayser and Christopher D. Manning 130

Decoder Integration and Expected BLEU Training for Recurrent Neural Network Language Models
Michael Auli and Jianfeng Gao 136

On the Elements of an Accurate Tree-to-String Machine Translation System
Graham Neubig and Kevin Duh 143

Simple extensions and POS Tags for a reparameterised IBM Model 2
Douwe Gelling and Trevor Cohn 150

Dependency-based Pre-ordering for Chinese-English Machine Translation
Jingsheng Cai, Masao Utiyama, Eiichiro Sumita and Yujie Zhang 155

Generalized Character-Level Spelling Error Correction
Noura Farra, Nadi Tomeh, Alla Rozovskaya and Nizar Habash 161

Improved Iterative Correction for Distant Spelling Errors
Sergey Gubanov, Irina Galinskaya and Alexey Baytin 168

Predicting Grammaticality on an Ordinal Scale
Michael Heilman, Aoife Cahill, Nitin Madnani, Melissa Lopez, Matthew Mulholland and Joel Tetreault 174

I'm a Belieber: Social Roles via Self-identification and Conceptual Attributes
Charley Beller, Rebecca Knowles, Craig Harman, Shane Bergsma, Margaret Mitchell and Benjamin Van Durme 181

Automatically Detecting Corresponding Edit-Turn-Pairs in Wikipedia
Johannes Daxenberger and Iryna Gurevych 187

Two Knives Cut Better Than One: Chinese Word Segmentation with Dual Decomposition
Mengqiu Wang, Rob Voigt and Christopher D. Manning 193

Effective Document-Level Features for Chinese Patent Word Segmentation
Si Li and Nianwen Xue 199

Word Segmentation of Informal Arabic with Domain Adaptation
Will Monroe, Spence Green and Christopher D. Manning 206

Resolving Lexical Ambiguity in Tensor Regression Models of Meaning
Dimitri Kartsaklis, Nal Kalchbrenner and Mehrnoosh Sadrzadeh 212

A Novel Content Enriching Model for Microblog Using News Corpus
Yunlun Yang, Zhihong Deng and Hongliang Yu 218

Learning Bilingual Word Representations by Marginalizing Alignments
Tomáš Kočiský, Karl Moritz Hermann and Phil Blunsom 224

Detecting Retries of Voice Search Queries
Rivka Levitan and David Elson 230

Sliding Alignment Windows for Real-Time Crowd Captioning
Mohammad Kazemi, Rahman Lavaee, Iftekhar Naim and Daniel Gildea 236

Detection of Topic and its Extrinsic Evaluation Through Multi-Document Summarization
Yoshimi Suzuki and Fumiyo Fukumoto 241

Content Importance Models for Scoring Writing From Sources
Beata Beigman Klebanov, Nitin Madnani, Jill Burstein and Swapna Somasundaran 247

Chinese Morphological Analysis with Character-level POS Tagging
Mo Shen, Hongxiao Liu, Daisuke Kawahara and Sadao Kurohashi 253

Part-of-Speech Tagging using Conditional Random Fields: Exploiting Sub-Label Dependencies for Improved Accuracy
Miikka Silfverberg, Teemu Ruokolainen, Krister Lindén and Mikko Kurimo 259

POS induction with distributional and morphological information using a distance-dependent Chinese restaurant process
Kairit Sirts, Jacob Eisenstein, Micha Elsner and Sharon Goldwater 265

Improving the Recognizability of Syntactic Relations Using Contextualized Examples
Aditi Muralidharan and Marti A. Hearst 272

How to Speak a Language without Knowing It
Xing Shi, Kevin Knight and Heng Ji 278

Assessing the Discourse Factors that Influence the Quality of Machine Translation
Junyi Jessy Li, Marine Carpuat and Ani Nenkova 283

Automatic Detection of Machine Translated Text and Translation Quality Estimation
Roee Aharoni, Moshe Koppel and Yoav Goldberg 289

Improving sparse word similarity models with asymmetric measures
Jean Mark Gawron 296

Dependency-Based Word Embeddings
Omer Levy and Yoav Goldberg 302

Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony
Florent Perek 309

Single Document Summarization based on Nested Tree Structure
Yuta Kikuchi, Tsutomu Hirao, Hiroya Takamura, Manabu Okumura and Masaaki Nagata 315

Linguistic Considerations in Automatic Question Generation
Karen Mazidi and Rodney D. Nielsen 321

Polynomial Time Joint Structural Inference for Sentence Compression
Xian Qian and Yang Liu 327

A Bayesian Method to Incorporate Background Knowledge during Automatic Text Summarization
Annie Louis 333

Predicting Power Relations between Participants in Written Dialog from a Single Thread
Vinodkumar Prabhakaran and Owen Rambow 339

Tri-Training for Authorship Attribution with Limited Training Data
Tieyun Qian, Bing Liu, Li Chen and Zhiyong Peng 345

Automation and Evaluation of the Keyword Method for Second Language Learning
Gözde Özbal, Daniele Pighin and Carlo Strapparava 352

Citation Resolution: A method for evaluating context-based citation recommendation systems
Daniel Duma and Ewan Klein 358

Hippocratic Abbreviation Expansion
Brian Roark and Richard Sproat 364

Unsupervised Feature Learning for Visual Sign Language Identification
Binyam Gebrekidan Gebre, Onno Crasborn, Peter Wittenburg, Sebastian Drude and Tom Heskes 370

Experiments with crowdsourced re-annotation of a POS tagging data set
Dirk Hovy, Barbara Plank and Anders Søgaard 377

Building Sentiment Lexicons for All Major Languages
Yanqing Chen and Steven Skiena 383

Difficult Cases: From Data to Learning, and Back
Beata Beigman Klebanov and Eyal Beigman 390

The VerbCorner Project: Findings from Phase 1 of crowd-sourcing a semantic decomposition of verbs
Joshua K. Hartshorne, Claire Bonial and Martha Palmer 397

A Corpus of Sentence-level Revisions in Academic Writing: A Step towards Understanding Statement Strength in Communication
Chenhao Tan and Lillian Lee 403

Determiner-Established Deixis to Communicative Artifacts in Pedagogical Text
Shomir Wilson and Jon Oberlander 409

Modeling Factuality Judgments in Social Media Text
Sandeep Soni, Tanushree Mitra, Eric Gilbert and Jacob Eisenstein 415

A Topic Model for Building Fine-grained Domain-specific Emotion Lexicon
Min Yang, Dingju Zhu and Kam-Pui Chow 421

Depeche Mood: a Lexicon for Emotion Analysis from Crowd Annotated News
Jacopo Staiano and Marco Guerini 427

Improving Twitter Sentiment Analysis with Topic-Based Mixture Modeling and Semi-Supervised Training
Bing Xiang and Liang Zhou 434

Cross-cultural Deception Detection
Verónica Pérez-Rosas and Rada Mihalcea 440

Particle Filter Rejuvenation and Latent Dirichlet Allocation
Chandler May, Alex Clemmer and Benjamin Van Durme 446

Comparing Automatic Evaluation Measures for Image Description
Desmond Elliott and Frank Keller 452

Learning a Lexical Simplifier Using Wikipedia
Colby Horn, Cathryn Manduca and David Kauchak 458

Cheap and easy entity evaluation
Ben Hachey, Joel Nothman and Will Radford 464

Identifying Real-Life Complex Task Names with Task-Intrinsic Entities from Microblogs
Ting-Xuan Wang, Kun-Yu Tsai and Wen-Hsiang Lu 470

Mutual Disambiguation for Entity Linking
Eric Charton, Marie-Jean Meurs, Ludovic Jean-Louis and Michel Gagnon 476

How Well can We Learn Interpretable Entity Types from Text?
Dirk Hovy 482

Learning Translational and Knowledge-based Similarities from Relevance Rankings for Cross-Language Retrieval
Shigehiko Schamoni, Felix Hieber, Artem Sokolov and Stefan Riezler 488

Two-Stage Hashing for Fast Document Retrieval
Hao Li, Wei Liu and Heng Ji 495

An Annotation Framework for Dense Event Ordering
Taylor Cassidy, Bill McDowell, Nathanael Chambers and Steven Bethard 501

Linguistically debatable or just plain wrong?
Barbara Plank, Dirk Hovy and Anders Søgaard 507

Humans Require Context to Infer Ironic Intent (so Computers Probably do, too)
Byron C. Wallace, Do Kook Choe, Laura Kertz and Eugene Charniak 512

Automatic prediction of aspectual class of verbs in context
Annemarie Friedrich and Alexis Palmer 517

Combining Word Patterns and Discourse Markers for Paradigmatic Relation Classification
Michael Roth and Sabine Schulte im Walde 524

Applying a Naive Bayes Similarity Measure to Word Sense Disambiguation
Tong Wang and Graeme Hirst 531

Fast Easy Unsupervised Domain Adaptation with Marginalized Structured Dropout
Yi Yang and Jacob Eisenstein 538

Improving Lexical Embeddings with Semantic Knowledge
Mo Yu and Mark Dredze 545

Optimizing Segmentation Strategies for Simultaneous Speech Translation
Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura 551

A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation
Taku Kudo, Hiroshi Ichikawa and Hideto Kazawa 557

A Hybrid Approach to Skeleton-based Translation
Tong Xiao, Jingbo Zhu and Chunliang Zhang 563

Effective Selection of Translation Model Training Data
Le Liu, Yu Hong, Hao Liu, Xing Wang and Jianmin Yao 569

Refinements to Interactive Translation Prediction Based on Search Graphs
Philipp Koehn, Chara Tsoukala and Herve Saint-Amand 574

Cross-lingual Model Transfer Using Feature Representation Projection
Mikhail Kozhevnikov and Ivan Titov 579

Cross-language and Cross-encyclopedia Article Linking Using Mixed-language Topic Model and Hypernym Translation
Yu-Chun Wang, Chun-Kai Wu and Richard Tzong-Han Tsai 586

Nonparametric Method for Data-driven Image Captioning
Rebecca Mason and Eugene Charniak 592

Improved Correction Detection in Revised ESL Sentences
Huichao Xue and Rebecca Hwa 599

Unsupervised Alignment of Privacy Policies using Hidden Markov Models
Rohan Ramanath, Fei Liu, Norman Sadeh and Noah A. Smith 605

Enriching Cold Start Personalized Language Model Using Social Network Information
Yu-Yang Huang, Rui Yan, Tsung-Ting Kuo and Shou-De Lin 611

Automatic Labelling of Topic Models Learned from Twitter by Summarisation
Amparo Elizabeth Cano Basave, Yulan He and Ruifeng Xu 618

Stochastic Contextual Edit Distance and Probabilistic FSTs
Ryan Cotterell, Nanyun Peng and Jason Eisner 625

Labelling Topics using Unsupervised Graph-based Methods
Nikolaos Aletras and Mark Stevenson 631

Training a Korean SRL System with Rich Morphological Features
Young-Bum Kim, Heemoon Chae, Benjamin Snyder and Yu-Seop Kim 637

Semantic Parsing for Single-Relation Question Answering
Wen-tau Yih, Xiaodong He and Christopher Meek 643

On WordNet Semantic Classes and Dependency Parsing
Kepa Bengoetxea, Eneko Agirre, Joakim Nivre, Yue Zhang and Koldo Gojenola 649

Enforcing Structural Diversity in Cube-pruned Dependency Parsing
Hao Zhang and Ryan McDonald 656

The Penn Parsed Corpus of Modern British English: First Parsing Results and Analysis
Seth Kulick, Anthony Kroch and Beatrice Santorini 662

Parser Evaluation Using Derivation Trees: A Complement to evalb
Seth Kulick, Ann Bies, Justin Mott, Anthony Kroch, Beatrice Santorini and Mark Liberman . . 668

Learning Polylingual Topic Models from Code-Switched Social Media Documents
Nanyun Peng, Yiming Wang and Mark Dredze 674

Normalizing tweets with edit scripts and recurrent neural embeddings
Grzegorz Chrupała 680

Exponential Reservoir Sampling for Streaming Language Models
Miles Osborne, Ashwin Lall and Benjamin Van Durme 687

A Piece of My Mind: A Sentiment Analysis Approach for Online Dispute Detection
Lu Wang and Claire Cardie 693

A Simple Bayesian Modelling Approach to Event Extraction from Twitter
Deyu Zhou, Liangyu Chen and Yulan He 700

Be Appropriate and Funny: Automatic Entity Morph Encoding
Boliang Zhang, Hongzhao Huang, Xiaoman Pan, Heng Ji, Kevin Knight, Zhen Wen, Yizhou Sun, Jiawei Han and Bulent Yener 706

Applying Grammar Induction to Text Mining
Andrew Salway and Samia Touileb 712

Semantic Consistency: A Local Subspace Based Method for Distant Supervised Relation Extraction
Xianpei Han and Le Sun 718

Concreteness and Subjectivity as Dimensions of Lexical Meaning
Felix Hill and Anna Korhonen 725

Infusion of Labeled Data into Distant Supervision for Relation Extraction
Maria Pershina, Bonan Min, Wei Xu and Ralph Grishman 732

Recognizing Implied Predicate-Argument Relationships in Textual Inference
Asher Stern and Ido Dagan 739

Measuring metaphoricity
Jonathan Dunn 745

Empirical Study of Unsupervised Chinese Word Segmentation Methods for SMT on Large-scale Corpora
Xiaolin Wang, Masao Utiyama, Andrew Finch and Eiichiro Sumita 752

EM Decipherment for Large Vocabularies
Malte Nuhn and Hermann Ney 759

XMEANT: Better semantic MT evaluation without reference translations
Chi-kiu Lo, Meriem Beloucif, Markus Saers and Dekai Wu 765

Sentence Level Dialect Identification for Machine Translation System Selection
Wael Salloum, Heba Elfardy, Linda Alamir-Salloum, Nizar Habash and Mona Diab 772

RNN-based Derivation Structure Prediction for SMT
Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong 779

Hierarchical MT Training using Max-Violation Perceptron
Kai Zhao, Liang Huang, Haitao Mi and Abe Ittycheriah 785

Punctuation Processing for Projective Dependency Parsing
Ji Ma, Yue Zhang and Jingbo Zhu 791

Transforming trees into hedges and parsing with "hedgebank" grammars
Mahsa Yarmohammadi, Aaron Dunlop and Brian Roark 797

Incremental Predictive Parsing with TurboParser
Arne Köhn and Wolfgang Menzel 803

Tailoring Continuous Word Representations for Dependency Parsing
Mohit Bansal, Kevin Gimpel and Karen Livescu 809

Observational Initialization of Type-Supervised Taggers
Hui Zhang and John DeNero 816

How much do word embeddings encode about syntax?
Jacob Andreas and Dan Klein 822

Distributed Representations of Geographically Situated Language
David Bamman, Chris Dyer and Noah A. Smith 828

Improving Multi-Modal Representations Using Image Dispersion: Why Less is Sometimes More
Douwe Kiela, Felix Hill, Anna Korhonen and Stephen Clark 835

Bilingual Event Extraction: a Case Study on Trigger Type Determination
Zhu Zhu, Shoushan Li, Guodong Zhou and Rui Xia 842

Understanding Relation Temporality of Entities
Taesung Lee and Seung-won Hwang 848

Does the Phonology of L1 Show Up in L2 Texts?
Garrett Nicolai and Grzegorz Kondrak 854

Cross-lingual Opinion Analysis via Negative Transfer Detection
Lin Gui, Ruifeng Xu, Qin Lu, Jun Xu, Jian Xu, Bin Liu and Xiaolong Wang 860

Conference Program

Sunday, June 22, 2014

7:30–18:00 Registration

7:30–9:00 Breakfast

9:00–12:30 Morning Tutorial

Session T1: Gaussian Processes for Natural Language Processing

Session T2: Scalable Large-Margin Structured Learning: Theory and Algorithms

Session T3: Semantics for Large-Scale Multimedia: New Challenges for NLP

Session T4: Wikification and Beyond: The Challenges of Entity and Concept Grounding

12:30–14:00 Lunch break

14:00–17:30 Afternoon Tutorial

Session T5: New Directions in Vector Space Models of Meaning

Session T6: Structured Belief Propagation for NLP

Session T7: Semantics, Discourse and Statistical Machine Translation

Session T8: Syntactic Processing Using Global Discriminative Learning and Beam-Search Decoding

18:00–21:00 Welcome Reception

Monday, June 23, 2014

7:30–18:00	Registration
7:30–9:00	Breakfast
8:55–9:00	Opening session
9:00–9:40	President talk
9:40–10:10	Coffee break
	Session 1A: Discourse, Dialogue, Coreference and Pragmatics
	Session 1B: Semantics I
	Session 1C: Machine Translation I
	Session 1D: Syntax, Parsing, and Tagging I
	Session 1E: NLP for the Web and Social Media I
11:50–13:20	Lunch break; Student Lunch

Monday, June 23, 2014 (continued)

Session 2A: Syntax, Parsing and Tagging II

Session 2B: Semantics II

Session 2C: Word Segmentation and POS Tagging

Session 2D: SRW

Session 2E: Sentiment Analysis I

15:00–15:30 Coffee break

Session 3A: Topic Modeling

Session 3B: Information Extraction I

Session 3C: Generation

Session 3D: Syntax, Parsing and Tagging III

Session 3E: Language Resources and Evaluation I

16:45–17:00 Break

17:00–18:00 Invited talk I: Corinna Cortes

Monday, June 23, 2014 (continued)

Oral Sessions for Student Research Workshop Posters

18:50–21:30 Poster and Dinner Session I: TACL Papers, Long Papers, Short Papers, Student Research Workshop; Demonstrations

Exploring the Relative Role of Bottom-up and Top-down Information in Phoneme Learning
Abdellah Fourtassi, Thomas Schatz, Balakrishnan Varadarajan and Emmanuel Dupoux

Biases in Predicting the Human Language Model
Alex B. Fine, Austin F. Frank, T. Florian Jaeger and Benjamin Van Durme

Probabilistic Labeling for Efficient Referential Grounding based on Collaborative Discourse
Changsong Liu, Lanbo She, Rui Fang and Joyce Y. Chai

A Composite Kernel Approach for Dialog Topic Tracking with Structured Domain Knowledge from Wikipedia
Seokhwan Kim, Rafael E. Banchs and Haizhou Li

An Extension of BLANC to System Mentions
Xiaoqiang Luo, Sameer Pradhan, Marta Recasens and Eduard Hovy

Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation
Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng and Michael Strube

Measuring Sentiment Annotation Complexity of Text
Aditya Joshi, Abhijit Mishra, Nivvedan Senthamilselvan and Pushpak Bhattacharyya

Improving Citation Polarity Classification with Product Reviews
Charles Jochim and Hinrich Schütze

Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification
Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou and Ke Xu

Sprinkling Topics for Weakly Supervised Text Classification
Swapnil Hingmire and Sutanu Chakraborti

A Feature-Enriched Tree Kernel for Relation Extraction
Le Sun and Xianpei Han

Employing Word Representations and Regularization for Domain Adaptation of Relation Extraction
Thien Huu Nguyen and Ralph Grishman

Graph Ranking for Collective Named Entity Disambiguation
Ayman Alhelbawy and Robert Gaizauskas

Descending-Path Convolution Kernel for Syntactic Structures
Chen Lin, Timothy Miller, Alvin Kho, Steven Bethard, Dmitriy Dligach, Sameer Pradhan and Guergana Savova

Entities' Sentiment Relevance
Zvi Ben-Ami, Ronen Feldman and Binyamin Rosenfeld

Automatic Detection of Multilingual Dictionaries on the Web
Gintare Grigonyte and Timothy Baldwin

Automatic Detection of Cognates Using Orthographic Alignment
Alina Maria Ciobanu and Liviu P. Dinu

Automatically constructing Wordnet Synsets
Khang Nhut Lam, Feras Al Tarouti and Jugal Kalita

Constructing a Turkish-English Parallel TreeBank
Olcay Taner Yıldız, Ercan Solak, Onur Görgün and Razieh Ehsani

Improved Typesetting Models for Historical OCR
Taylor Berg-Kirkpatrick and Dan Klein

Robust Logistic Regression using Shift Parameters
Julie Tibshirani and Christopher D. Manning

Faster Phrase-Based Decoding by Refining Feature State
Kenneth Heafield, Michael Kayser and Christopher D. Manning

Decoder Integration and Expected BLEU Training for Recurrent Neural Network Language Models
Michael Auli and Jianfeng Gao

Monday, June 23, 2014 (continued)

On the Elements of an Accurate Tree-to-String Machine Translation System
Graham Neubig and Kevin Duh

Simple extensions and POS Tags for a reparameterised IBM Model 2
Douwe Gelling and Trevor Cohn

Dependency-based Pre-ordering for Chinese-English Machine Translation
Jingsheng Cai, Masao Utiyama, Eiichiro Sumita and Yujie Zhang

Generalized Character-Level Spelling Error Correction
Noura Farra, Nadi Tomeh, Alla Rozovskaya and Nizar Habash

Improved Iterative Correction for Distant Spelling Errors
Sergey Gubanov, Irina Galinskaya and Alexey Baytin

Predicting Grammaticality on an Ordinal Scale
Michael Heilman, Aoife Cahill, Nitin Madnani, Melissa Lopez, Matthew Mulholland and Joel Tetreault

I'm a Belieber: Social Roles via Self-identification and Conceptual Attributes
Charley Beller, Rebecca Knowles, Craig Harman, Shane Bergsma, Margaret Mitchell and Benjamin Van Durme

Automatically Detecting Corresponding Edit-Turn-Pairs in Wikipedia
Johannes Daxenberger and Iryna Gurevych

Two Knives Cut Better Than One: Chinese Word Segmentation with Dual Decomposition
Mengqiu Wang, Rob Voigt and Christopher D. Manning

Effective Document-Level Features for Chinese Patent Word Segmentation
Si Li and Nianwen Xue

Word Segmentation of Informal Arabic with Domain Adaptation
Will Monroe, Spence Green and Christopher D. Manning

Resolving Lexical Ambiguity in Tensor Regression Models of Meaning
Dimitri Kartsaklis, Nal Kalchbrenner and Mehrnoosh Sadrzadeh

A Novel Content Enriching Model for Microblog Using News Corpus
Yunlun Yang, Zhihong Deng and Hongliang Yu

Learning Bilingual Word Representations by Marginalizing Alignments
Tomáš Kočiský, Karl Moritz Hermann and Phil Blunsom

Detecting Retries of Voice Search Queries
Rivka Levitan and David Elson

Sliding Alignment Windows for Real-Time Crowd Captioning
Mohammad Kazemi, Rahman Lavaee, Iftekhar Naim and Daniel Gildea

Detection of Topic and its Extrinsic Evaluation Through Multi-Document Summarization
Yoshimi Suzuki and Fumiyo Fukumoto

Content Importance Models for Scoring Writing From Sources
Beata Beigman Klebanov, Nitin Madnani, Jill Burstein and Swapna Somasundaran

Chinese Morphological Analysis with Character-level POS Tagging
Mo Shen, Hongxiao Liu, Daisuke Kawahara and Sadao Kurohashi

Part-of-Speech Tagging using Conditional Random Fields: Exploiting Sub-Label Dependencies for Improved Accuracy
Miikka Silfverberg, Teemu Ruokolainen, Krister Lindén and Mikko Kurimo

POS induction with distributional and morphological information using a distance-dependent Chinese restaurant process
Kairit Sirts, Jacob Eisenstein, Micha Elsner and Sharon Goldwater

Improving the Recognizability of Syntactic Relations Using Contextualized Examples
Aditi Muralidharan and Marti A. Hearst

Tuesday, June 24, 2014

7:30–18:00 Registration

7:30–9:00 Breakfast

9:00–10:00 Invited talk II: Zoran Popović

10:00–10:30 Coffee break

Session 4A: Machine Learning for NLP

Session 4B: Information Extraction II

Session 4C: Machine Translation II

Session 4D: Summarization

Session 4E: Language Resources and Evaluation II

12:10–13:30 Lunch break

Session 5A: Question Answering

Session 5B: Information Extraction III

Tuesday, June 24, 2014 (continued)

Session 5C: Lexical Semantics and Ontology I

Session 5D: Syntax, Parsing and Tagging IV

Session 5E: Cognitive Modeling and Psycholinguistics

14:45–15:15 Coffee break

Session 6A: Machine Translation III

15:15–15:30 *How to Speak a Language without Knowing It*
Xing Shi, Kevin Knight and Heng Ji

15:30–15:45 *Assessing the Discourse Factors that Influence the Quality of Machine Translation*
Junyi Jessy Li, Marine Carpuat and Ani Nenkova

15:45–16:00 *Automatic Detection of Machine Translated Text and Translation Quality Estimation*
Roee Aharoni, Moshe Koppel and Yoav Goldberg

Session 6B: Lexical Semantics and Ontology II

15:15–15:30 *Improving sparse word similarity models with asymmetric measures*
Jean Mark Gawron

15:30–15:45 *Dependency-Based Word Embeddings*
Omer Levy and Yoav Goldberg

15:45–16:00 *Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony*
Florent Perek

Tuesday, June 24, 2014 (continued)

Session 6C: Generation/Summarization/Dialogue

15:15–15:30 *Single Document Summarization based on Nested Tree Structure*
Yuta Kikuchi, Tsutomu Hirao, Hiroya Takamura, Manabu Okumura and Masaaki Nagata

15:30–15:45 *Linguistic Considerations in Automatic Question Generation*
Karen Mazidi and Rodney D. Nielsen

15:45–16:00 *Polynomial Time Joint Structural Inference for Sentence Compression*
Xian Qian and Yang Liu

16:00–16:15 *A Bayesian Method to Incorporate Background Knowledge during Automatic Text Summarization*
Annie Louis

16:15–16:30 *Predicting Power Relations between Participants in Written Dialog from a Single Thread*
Vinodkumar Prabhakaran and Owen Rambow

Session 6D: NLP Applications and NLP Enabled Technology I

15:15–15:30 *Tri-Training for Authorship Attribution with Limited Training Data*
Tieyun Qian, Bing Liu, Li Chen and Zhiyong Peng

15:30–15:45 *Automation and Evaluation of the Keyword Method for Second Language Learning*
Gözde Özbal, Daniele Pighin and Carlo Strapparava

15:45–16:00 *Citation Resolution: A method for evaluating context-based citation recommendation systems*
Daniel Duma and Ewan Klein

16:00–16:15 *Hippocratic Abbreviation Expansion*
Brian Roark and Richard Sproat

16:15–16:30 *Unsupervised Feature Learning for Visual Sign Language Identification*
Binyam Gebrekidan Gebre, Onno Crasborn, Peter Wittenburg, Sebastian Drude and Tom Heskes

Tuesday, June 24, 2014 (continued)

Session 6E: Language Resources and Evaluation III

15:15–15:30 *Experiments with crowdsourced re-annotation of a POS tagging data set*
Dirk Hovy, Barbara Plank and Anders Søgaard

15:30–15:45 *Building Sentiment Lexicons for All Major Languages*
Yanqing Chen and Steven Skiena

15:45–16:00 *Difficult Cases: From Data to Learning, and Back*
Beata Beigman Klebanov and Eyal Beigman

16:00–16:15 *The VerbCorner Project: Findings from Phase 1 of crowd-sourcing a semantic decomposition of verbs*
Joshua K. Hartshorne, Claire Bonial and Martha Palmer

16:15–16:30 *A Corpus of Sentence-level Revisions in Academic Writing: A Step towards Understanding Statement Strength in Communication*
Chenhao Tan and Lillian Lee

16:50–19:20 Poster and Dinner Session II: Long Papers, Short Papers and Demonstrations in Grand BallroomI/II/III/IV/V/VI/VII/VIII

Determiner-Established Deixis to Communicative Artifacts in Pedagogical Text
Shomir Wilson and Jon Oberlander

Modeling Factuality Judgments in Social Media Text
Sandeep Soni, Tanushree Mitra, Eric Gilbert and Jacob Eisenstein

A Topic Model for Building Fine-grained Domain-specific Emotion Lexicon
Min Yang, Dingju Zhu and Kam-Pui Chow

Depeche Mood: a Lexicon for Emotion Analysis from Crowd Annotated News
Jacopo Staiano and Marco Guerini

Improving Twitter Sentiment Analysis with Topic-Based Mixture Modeling and Semi-Supervised Training
Bing Xiang and Liang Zhou

Cross-cultural Deception Detection
Verónica Pérez-Rosas and Rada Mihalcea

Particle Filter Rejuvenation and Latent Dirichlet Allocation
Chandler May, Alex Clemmer and Benjamin Van Durme

Comparing Automatic Evaluation Measures for Image Description
Desmond Elliott and Frank Keller

Learning a Lexical Simplifier Using Wikipedia
Colby Horn, Cathryn Manduca and David Kauchak

Cheap and easy entity evaluation
Ben Hachey, Joel Nothman and Will Radford

Identifying Real-Life Complex Task Names with Task-Intrinsic Entities from Microblogs
Ting-Xuan Wang, Kun-Yu Tsai and Wen-Hsiang Lu

Mutual Disambiguation for Entity Linking
Eric Charton, Marie-Jean Meurs, Ludovic Jean-Louis and Michel Gagnon

How Well can We Learn Interpretable Entity Types from Text?
Dirk Hovy

Learning Translational and Knowledge-based Similarities from Relevance Rankings for Cross-Language Retrieval
Shigehiko Schamoni, Felix Hieber, Artem Sokolov and Stefan Riezler

Two-Stage Hashing for Fast Document Retrieval
Hao Li, Wei Liu and Heng Ji

An Annotation Framework for Dense Event Ordering
Taylor Cassidy, Bill McDowell, Nathanael Chambers and Steven Bethard

Linguistically debatable or just plain wrong?
Barbara Plank, Dirk Hovy and Anders Søgaard

Humans Require Context to Infer Ironic Intent (so Computers Probably do, too)
Byron C. Wallace, Do Kook Choe, Laura Kertz and Eugene Charniak

Automatic prediction of aspectual class of verbs in context
Annemarie Friedrich and Alexis Palmer

Combining Word Patterns and Discourse Markers for Paradigmatic Relation Classification
Michael Roth and Sabine Schulte im Walde

Applying a Naive Bayes Similarity Measure to Word Sense Disambiguation
Tong Wang and Graeme Hirst

Fast Easy Unsupervised Domain Adaptation with Marginalized Structured Dropout
Yi Yang and Jacob Eisenstein

Improving Lexical Embeddings with Semantic Knowledge
Mo Yu and Mark Dredze

Optimizing Segmentation Strategies for Simultaneous Speech Translation
Yusuke Oda, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura

A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation
Taku Kudo, Hiroshi Ichikawa and Hideto Kazawa

A Hybrid Approach to Skeleton-based Translation
Tong Xiao, Jingbo Zhu and Chunliang Zhang

Effective Selection of Translation Model Training Data
Le Liu, Yu Hong, Hao Liu, Xing Wang and Jianmin Yao

Refinements to Interactive Translation Prediction Based on Search Graphs
Philipp Koehn, Chara Tsoukala and Herve Saint-Amand

Cross-lingual Model Transfer Using Feature Representation Projection
Mikhail Kozhevnikov and Ivan Titov

Cross-language and Cross-encyclopedia Article Linking Using Mixed-language Topic Model and Hypernym Translation
Yu-Chun Wang, Chun-Kai Wu and Richard Tzong-Han Tsai

Nonparametric Method for Data-driven Image Captioning
Rebecca Mason and Eugene Charniak

Improved Correction Detection in Revised ESL Sentences
Huichao Xue and Rebecca Hwa

Unsupervised Alignment of Privacy Policies using Hidden Markov Models
Rohan Ramanath, Fei Liu, Norman Sadeh and Noah A. Smith

Enriching Cold Start Personalized Language Model Using Social Network Information
Yu-Yang Huang, Rui Yan, Tsung-Ting Kuo and Shou-De Lin

Automatic Labelling of Topic Models Learned from Twitter by Summarisation
Amparo Elizabeth Cano Basave, Yulan He and Ruifeng Xu

Stochastic Contextual Edit Distance and Probabilistic FSTs
Ryan Cotterell, Nanyun Peng and Jason Eisner

Labelling Topics using Unsupervised Graph-based Methods
Nikolaos Aletras and Mark Stevenson

Training a Korean SRL System with Rich Morphological Features
Young-Bum Kim, Heemoon Chae, Benjamin Snyder and Yu-Seop Kim

Semantic Parsing for Single-Relation Question Answering
Wen-tau Yih, Xiaodong He and Christopher Meek

On WordNet Semantic Classes and Dependency Parsing
Kepa Bengoetxea, Eneko Agirre, Joakim Nivre, Yue Zhang and Koldo Gojenola

Enforcing Structural Diversity in Cube-pruned Dependency Parsing
Hao Zhang and Ryan McDonald

The Penn Parsed Corpus of Modern British English: First Parsing Results and Analysis
Seth Kulick, Anthony Kroch and Beatrice Santorini

Tuesday, June 24, 2014 (continued)

Parser Evaluation Using Derivation Trees: A Complement to evalb
Seth Kulick, Ann Bies, Justin Mott, Anthony Kroch, Beatrice Santorini and Mark Liberman

19:30–22:00 Social at the National Aquarium in Baltimore

Wednesday, June 25, 2014

7:30–18:00 Registration

7:30–9:00 Breakfast

Best paper session

10:15–10:45 Coffee break

Session 7A: Multimodal NLP/ Lexical Semantics

Session 7B: Semantics III

Session 7C: Machine Translation IV

Session 7D: NLP Applications and NLP Enabled Technology II

Session 7E: Sentiment Analysis II

12:25–13:30 Lunch break

13:30–15:00 ACL Business Meeting

Wednesday, June 25, 2014 (continued)

Session 8A: NLP for the Web and Social Media II

15:00–15:15 *Learning Polylingual Topic Models from Code-Switched Social Media Documents*
Nanyun Peng, Yiming Wang and Mark Dredze

15:15–15:30 *Normalizing tweets with edit scripts and recurrent neural embeddings*
Grzegorz Chrupała

15:30–15:45 *Exponential Reservoir Sampling for Streaming Language Models*
Miles Osborne, Ashwin Lall and Benjamin Van Durme

15:45–16:00 *A Piece of My Mind: A Sentiment Analysis Approach for Online Dispute Detection*
Lu Wang and Claire Cardie

16:00–16:15 *A Simple Bayesian Modelling Approach to Event Extraction from Twitter*
Deyu Zhou, Liangyu Chen and Yulan He

16:15–16:30 *Be Appropriate and Funny: Automatic Entity Morph Encoding*
Boliang Zhang, Hongzhao Huang, Xiaoman Pan, Heng Ji, Kevin Knight, Zhen Wen, Yizhou Sun, Jiawei Han and Bulent Yener

Session 8B: Semantics/Information Extraction

15:00–15:15 *Applying Grammar Induction to Text Mining*
Andrew Salway and Samia Touileb

15:15–15:30 *Semantic Consistency: A Local Subspace Based Method for Distant Supervised Relation Extraction*
Xianpei Han and Le Sun

15:30–15:45 *Concreteness and Subjectivity as Dimensions of Lexical Meaning*
Felix Hill and Anna Korhonen

15:45–16:00 *Infusion of Labeled Data into Distant Supervision for Relation Extraction*
Maria Pershina, Bonan Min, Wei Xu and Ralph Grishman

16:00–16:15 *Recognizing Implied Predicate-Argument Relationships in Textual Inference*
Asher Stern and Ido Dagan

Wednesday, June 25, 2014 (continued)

16:15–16:30 *Measuring metaphoricity*
Jonathan Dunn

Session 8C: Machine Translation V

15:00–15:15 *Empirical Study of Unsupervised Chinese Word Segmentation Methods for SMT on Large-scale Corpora*
Xiaolin Wang, Masao Utiyama, Andrew Finch and Eiichiro Sumita

15:15–15:30 *EM Decipherment for Large Vocabularies*
Malte Nuhn and Hermann Ney

15:30–15:45 *XMEANT: Better semantic MT evaluation without reference translations*
Chi-kiu Lo, Meriem Beloucif, Markus Saers and Dekai Wu

15:45–16:00 *Sentence Level Dialect Identification for Machine Translation System Selection*
Wael Salloum, Heba Elfardy, Linda Alamir-Salloum, Nizar Habash and Mona Diab

16:00–16:15 *RNN-based Derivation Structure Prediction for SMT*
Feifei Zhai, Jiajun Zhang, Yu Zhou and Chengqing Zong

16:15–16:30 *Hierarchical MT Training using Max-Violation Perceptron*
Kai Zhao, Liang Huang, Haitao Mi and Abe Ittycheriah

Session 8D: Syntax, Parsing, and Tagging V

15:00–15:15 *Punctuation Processing for Projective Dependency Parsing*
Ji Ma, Yue Zhang and Jingbo Zhu

15:15–15:30 *Transforming trees into hedges and parsing with "hedgebank" grammars*
Mahsa Yarmohammadi, Aaron Dunlop and Brian Roark

15:30–15:45 *Incremental Predictive Parsing with TurboParser*
Arne Köhn and Wolfgang Menzel

15:45–16:00 *Tailoring Continuous Word Representations for Dependency Parsing*
Mohit Bansal, Kevin Gimpel and Karen Livescu

Wednesday, June 25, 2014 (continued)

16:00–16:15 *Observational Initialization of Type-Supervised Taggers*
Hui Zhang and John DeNero

16:15–16:30 *How much do word embeddings encode about syntax?*
Jacob Andreas and Dan Klein

Session 8E: Multilinguality and Multimodal NLP

15:00–15:15 *Distributed Representations of Geographically Situated Language*
David Bamman, Chris Dyer and Noah A. Smith

15:15–15:30 *Improving Multi-Modal Representations Using Image Dispersion: Why Less is Sometimes More*
Douwe Kiela, Felix Hill, Anna Korhonen and Stephen Clark

15:30–15:45 *Bilingual Event Extraction: a Case Study on Trigger Type Determination*
Zhu Zhu, Shoushan Li, Guodong Zhou and Rui Xia

15:45–16:00 *Understanding Relation Temporality of Entities*
Taesung Lee and Seung-won Hwang

16:00–16:15 *Does the Phonology of L1 Show Up in L2 Texts?*
Garrett Nicolai and Grzegorz Kondrak

16:15–16:30 *Cross-lingual Opinion Analysis via Negative Transfer Detection*
Lin Gui, Ruifeng Xu, Qin Lu, Jun Xu, Jian Xu, Bin Liu and Xiaolong Wang

16:30–17:00 Coffee break

17:00–18:30 Lifetime Achievement Award

18:30–19:00 Closing session